Rule Training for VMI Sketch in Developmental Testing based on a Deep Neural Network

Tae-Gyun Lee[†‡] **and Jang-Hee Yoo**[†‡]
[†]ETRI - Artificial Intelligence Laboratory
[‡]University of Science and Technology (UST)
Daejeon 34129, South Korea
{tae0086|jhy}@etri.re.kr

Abstract

In this paper, we present a framework that explains the scores of sketches by learning rules used in developmental tests. To achieve this, we propose a deep neural network model that considers a target and the corresponding sketch images as inputs. The proposed method is divided into plain and residual models according to the presence of residual connections to compare their performance. In addition, each model includes the subtraction and concatenation approaches to fuse two feature maps. To verify the performance of the proposed method, we conduct experiments over all settings that combine the proposed models with the fusion method. The results show that the proposed framework can be used for visual-motor integration analysis by determining scores and providing explanations.

1 Introduction

Developmental tests are used to evaluate a child's abilities and compare them with those of others of a similar age [1]. Given that early intervention yields a better prognosis in many developmental disorders, it is important to check children's developmental levels early through developmental tests. The types of developmental disabilities that can be diagnosed through developmental tests include intellectual disabilities, autism spectrum disorder (ASD), and attention-deficit/hyperactivity disorder (ADHD). In general, infants between 12 and 24 months of age who do not know how to write naturally have difficulty in performing developmental tests using writing. Instead, developmental tests can be performed using sketches. Some examples of such tests include the Rey–Osterrieth complex figure (ROCF) [2] and visual-motor integration (VMI) [3]. Moreover, VMI can be an important variable in handwriting, and Sanghavi and Kelkar [4] found it to be the most important factor in handwriting performance.

VMI refers to the ability to coordinate one's eyes and hands together smoothly and efficiently [4]. Measures of VMI skills are widely used to evaluate the mental and psychological state of children [5, 6]. The most widespread tool for VMI assessment is the Beery-Buktenica developmental test of visual-motor integration, known as the Beery VMI [6]. It uses sketches that consist of diverse figures. The figures used in the Beery VMI range from simple forms such as a line and a circle to complex designs such as a cube and a three-dimensional star. The test is performed using test sheets that comprise of six boundaries, and the figures are displayed in the upper row. The children see the figures and draw them on the corresponding spaces below. Subsequently, experts inspect the drawn sketches only in the boundaries and score each sketch with 0 or 1 points according to the criteria (rules) provided in the Beery VMI scoring manual [6].

Studies have been conducted with the Beery VMI to assess specific groups in terms of ethnicity, disorder, age, etc. Specifically, Coallier et al. [7] used the test for kindergarten children. Green et al. [8] compared the VMI abilities of children with ASD with typical development groups. Geldof et al.

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

[9] verified the deficit of VMI ability in children born prematurely. On the other hand, some studies have used automated methods to recognize the forms in a sketch or retrieve images using them. Yu et al. [10] and Zhang et al. [11] used deep neural networks (DNNs) to perform sketch recognition. Wang et al. [12] and Sain et al. [13] studied a sketch-based image retrieval method that used sketches as input to find similar images. Although some previous works have considered the Beery VMI and sketches, few studies have used DNNs to describe the sketches collected in the Beery VMI.

In this paper, we propose a framework called a sketch rule-training network (SRTN) to explain scoring results in the VMI analysis. From a clinical perspective, the explanation of the scoring result is more important than the result itself. The reason for a given score can be inferred based on the rules because the Beery VMI has figure-specific explicit rules. However, the score of any given figure may differ depending on subjective judgements of experts. To solve the issue, a rule inference algorithm that reflects such empirical cases is required. To incorporate diverse rules into single network, we use target figures as input data in addition to the sketches.

2 Proposed Method

The proposed rule-training method for sketch explanation requires two images, including a target figure and a sketch image. The VMI data used includes obvious rules, which the proposed neural network learns to enable the model to infer explanations of the input sketch. The overall architecture of the proposed framework is demonstrated in Figure 1. For rule training, the target figure images are also required along with the sketch images. As a network outputs the rule codes, they must be decoded to obtain the text explanation.

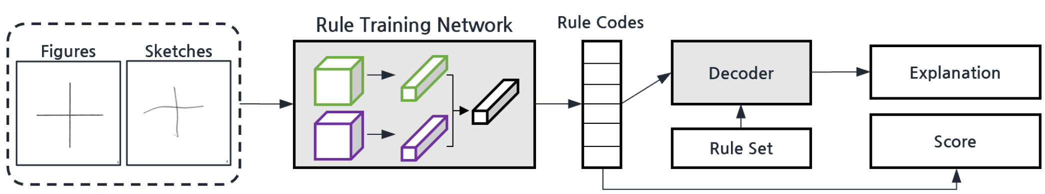


Figure 1: The proposed architecture of the SRTN framework.

We used a convolutional neural network (CNN) to learn them with a single integrated model because the types of figures and the scoring rules used in this study were diverse. The overall architecture of the proposed model is shown in Table 1. The used neural network is largely divided into a plain model using a general convolutional layer and a residual model using a residual block such as VGG and ResNet, respectively. The output nodes of the proposed network are set to 6, which is the maximum number of criteria for all figure items. In addition, 1 and -1 outputs represent 1 and 0 points respectively, and when there were fewer than 6 criteria, the output of the remaining nodes was set to 0. The explanation results were decoded by the inference and corresponding rules.

As shown in Table 1, one input is a sketch, while the other is a target figure representing the items shown in the sketch. The figure images are input together because the neural network can extract features depending on which figure form the sketch exhibits. Consequently, information can be learned from various figures with a single integrated neural network. The input sketch and figure image pass through two independent networks to output two feature maps. The output feature maps

Table 1: Overall architecture of the DNN model used within this study.

Name	Output Size	Plain or Residual	
		Target figure	Sketch
stem	112×112 56×56	7×7, 64, stride 2, padding 3 2×2 max pool, stride 2	7×7, 64, stride 2, padding 3 2×2 max pool, stride 2
conv block	14×14	plain or residual block x2	plain or residual block x2
fusion	14×14	subtraction or concatenation	
conv	14×14	1×1, 256	
output	6	average pool, 255-d, fc, tanh	

are combined into a single map to predict whether the scoring criterion is satisfied. Two approaches are used to combine the feature maps. The first is a subtraction that calculates the difference between the two feature maps, while the other is concatenation that attaches all information without loss.

After the trained DNN inference with the input images, the explanation is generated by decoding the result. A rule set was determined from predefined sets of VMI figures according to the target figure as a neural network input, and the rules are combined with the inference codes in a way that depends on whether the corresponding rules are satisfied. Since the number and order of rules in each figure are predefined, the decoder maps the encoded output directly to the learned rules. When it is determined whether the rule is satisfied, the explanation is completed by combining decoded rules with appropriate phrases. Moreover, the score that indicates whether all rules in the figure are satisfied is also decoded. According to the VMI manual, the score for a figure is 1 when all the criteria are satisfied, and the score is 0 otherwise. The examples of the rules on two test figures are displayed in Table 2.

Table 2: The rules for two sample figures in the Beery VMI test [6].

Figure	Rule
	· The line is within 30° horizontally
	· The length of overlapping/gap of open squares and circles is less than 1.6mm · No serious distortions · The ratio of the size of the circle to the open square is similar (within 2:1) · The line that bisects the circle is within the open square

3 Experimental Results

The sketch data collected included twelve types from the Beery VMI test, but only six types of sketches were used in the experiment. A total of 133 test sheets were used, and the number of individual figures was 399 because a single sheet consists of three figures. To reserve additional training data, the target figures and sketch examples presented in the VMI manual were additionally collected. Furthermore, various data augmentation methods such as image re-scaling were used to increase the training data. To train and perform inferences only on a fixed area of the test sheet, the target figure and sketch areas were detected and segmented. The individual figure and sketch images were saved with a resolution of 1,000 × 1,000 pixels and then resized to 224 × 224 pixels before being used in the DNN models. Table 3 shows the number of data samples used in the experiment.

Table 3: The number of experimental sketch data samples used.

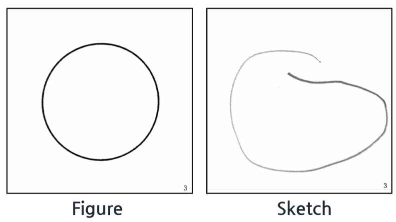
Categories		Count	
		6 figures	12 figures
Training	VMI tests	337	372
	VMI manual	91	188
Test	VMI tests	62	66
Total		492	626

Table 4 shows the performance results of all DNN models. In the plain model, the concatenation approach exhibited a n accuracy of 79.03%, which was 3.22% higher than the accuracy of the subtraction approach at 75.81%. Also, the accuracy of the subtraction was 80.65%, and the concatenation approach showed an accuracy of 82.26% in the residual model, which was 1.61% higher. Additionally, the performance over 12 types of figures was measured together for comparison with 6 types. As expected, training all collected data adversely affected the performance due to the lack of sufficient data.

Table 4: Performance results for DNN models.

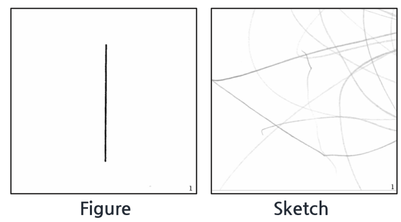
Model		Accuracy(%)	
		6 figures	12 figures
Plain	Subtraction	75.81	71.21
	Concatenation	79.03	74.24
Residual	Subtraction	80.65	71.21
	Concatenation	82.26	69.70

In the plain model, the accuracy of the subtraction was 71.21% and the accuracy of the concatenation was 74.24%, i.e., the accuracy of the concatenation approach was about 3.03% higher than the subtraction approach. In the residual model, the accuracy of the subtraction and concatenation approaches was 71.21% and 69.70%, respectively. In this model, the subtraction approach performed 1.51% better than the concatenation approach.



- Prediction, Target: Since it is satisfied that the ratio of the height to the width of the circle is within 2:1, the score is 1.

(a) a circle



- Prediction: Since it is satisfied that at least half of the line deviates more than 30° from the vertical, the score is 0.
- Target: Since it is satisfied that at least half of the line does not deviate more than 30° from the vertical, the score is 1.

(b) a vertical line

Figure 2: Decoding results on VMI test sketches. (a) received a score of 1 because the prediction and target explanation were the same, but (b) received a score of 0 because they were not.

As a result, Figure 2 demonstrates examples of explanations generated by the decoding inference output through the mapping to scoring rules according to the individual target figures. The prediction and the label match in Figure 4(a), whereas they do not match in Figure 4(b). The decoding results indicate that the DNN, which was trained on the rules in the proposed method, was able to explain the VMI sketches.

Since the experimental data were collected from actual developmental tests with less than 48 months infants, most of the sketches correspond to the first six types of figures. If more data are collected and extended sketches to other developmental tests, it is expected that the proposed method can be contributed to the early screening of developmental disorders.

4 Conclusions

We have described an automated scoring method and an associated explanation-generating method designed to learn rules using a deep neural network to measure visual-motor coordination ability using the Beery VMI sketch data. The DNN was designed as a plain model and a residual model, and the subtraction and concatenation approaches were used as a fusion of feature maps in each model. The results of an experiment verified that the architecture of the proposed model can be used to explain the scoring results. It showed a good performance of up to 82.26% accuracy by decoding the rules, although there were differences in performance depending on the method used. In future study, we plan to explore contrastive learning to improve the performance of explanations with better sketch representations. We expect that this will result in a performance improvement. Also, to generalize the rule training, utilization of a finite state machine on the rule set could also be considered.

Acknowledgments

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).

References

[1] Bellman, M., Byrne, O. & Sege, R. (2013) *Developmental Assessment of Children*, Bmj, **346**.

[2] Canham, R. O., Smith, S. L. & Tyrrell, A. M. (2000) Automated Scoring of A Neuropsychological Test: The Rey Osterrieth Complex Figure. In *Proceedings of the 26th Euromicro Conference. EUROMICRO 2000. Informatics: Inventing the Future*, **2**, pp.406-413.

[3] Hoque, S., Fairhurst, M. C. & Razian, M. A. (2004) Modulating Population Granularity for Improved Diagnosis of Developmental Dyspraxia from Dynamic Drawing Analysis. In *Proceedings of the Ninth International Workshop on Frontiers in Handwriting Recognition*, pp.26-31.

[4] Sanghavi, R. & Kelkar, R. (2005) Visual-Motor Integration and Learning Disabled Children. *The Indian Journal of Occupational Therapy* **37**(2), pp.33-38.

[5] Decker, S. L., Englund, J. A., Carboni, J. A. & Brooks, J. H. (2011) Cognitive and Developmental Influences in Visual-Motor Integration Skills in Young Children. *Psychological Assessment*, **23**(4), pp.1010.

[6] Beery, K. E. & Beery, N. A. (2010) *The Beery-Buktenica Developmental Test of Visual-Motor Integration (Beery-VMI) with Supplemental Developmental Tests of Visual Perception and Motor Coordination and Stepping Stones Age Norms: Administration, Scoring and Teaching Manual Sixth Edition*. Minneapolis, MN: NCS Pearson.

[7] Coallier, M., Rouleau, N., Bara, F. & Morin, M. F. (2014) Visual-Motor Skills Performance on the Beery-VMI: A Study of Canadian Kindergarten Children. *The Open Journal of Occupational Therapy*, **2**(2), pp.4.

[8] Green, R. R., Bigler, E. D., Froehlich, A., Prigge, M. B., Travers, B. G. et al. (2016) Beery VMI Performance in Autism Spectrum Disorder. *Child Neuropsychology*, **22**(7), pp.795-817.

[9] Geldof, C. J. A., Van Wassenaer, A. G., De Kieviet, J. F., Kok, J. H., & Oosterlaan, J. (2012) Visual Perception and Visual-Motor Integration in Very Preterm and/or Very Low Birth Weight Children: A Meta-Analysis. *Research in Developmental Disabilities* **33**(2), pp.726-736.

[10] Yu, Q., Yang, Y., Liu, F., Song, Y. Z., Xiang, T. et al. (2017) Sketch-A-Net: A Deep Neural Network That Beats Humans. *International Journal of Computer Vision*, **122**(3), pp.411-425.

[11] Zhang, X., Huang, Y., Zou, Q., Pei, Y., Zhang, R. et al. (2020) A Hybrid Convolutional Neural Network for Sketch Recognition. *Pattern Recognition Letters*, **130**, pp.73-82.

[12] Wang, X., Duan, X. & Bai, X. (2016) Deep Sketch Feature for Cross-Domain Image Retrieval. *Neurocomputing*, **207**, pp.387-397.

[13] Sain, A., Bhunia, A. K., Yang, Y., Xiang, T. & Song, Y. Z. (2021) Stylemeup: Towards Style-Agnostic Sketch-Based Image Retrieval. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp.8504-8513.