Recognition of Hand-drawn images of infants using Convolutional Neural Networks

Mi-Hwa Song[a]

[a] Assistant Professor of School of Information and Communication Science, Semyung University, Republic of Korea

_____

Abstract: The recognition rate of children's paintings is significantly lower than that of adults' paintings due to their unique characteristics. According to the research on children's art, children's paintings are self-centered and have many features different from those of adults, such as a lot of exaggerated expressions. In this paper, we introduce a method to increase the recognition rate of these children's pictures using deep learning. In order to improve the low recognition rate of children's pictures, a pre-processor that generalizes children's unique features was created to primarily purify the data, and classified into 250 using the Convolutional Neural Network, which is actively researched in the field of image recognition. High accuracy was obtained as a result of securing and executing sketch drawings of 80 adults for each object. Through this study, it is expected that not only improving the recognition ability for infant pictures, but also measuring learning ability and child development through children's drawings, and child psychotherapy through emotional recognition.

Keywords: Convolutional Neural Networks, Deep Learning, Hand-drawn, Pattern Recognition, Children's drawing

_____

1. Introduction

One of human's basic desires, the desire for expression, can be found in various art activities. Among them, free-hand sketch is a method that humanity has used for a long time as a universal means of expression of opinion that is freely used without using tools such as rulers, compasses, and paints.

In addition, due to the recent development of artificial intelligence technology, image recognition and processing technology has resulted in human art activities being systematically recognized and usable. However, unlike the freehand sketches of adults, children's drawings are still at a remarkably low level of recognition. This is because the characteristics of children's drawings are different from those of adults and appear in very various forms, so it is difficult to find a clear pattern.
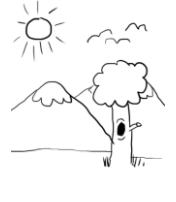
In this study, the characteristics of children's art studied in early childhood art and art education are classified in order to increase the recognition rate of these children's paintings. And, under the assumption that the sketch data of adult drawings contains children's drawings, deep learning techniques are applied to evaluate the results of schech image recognition.

2. Related Works

2.1 Features of Children's drawings

In general, a picture of an infant, unlike an image of an adult, expresses one's thoughts through a unique shape, which can be difficult to understand from an adult's point of view. From the perspective of infant art, Lowenfeld divides the main characteristics of infants according to their developmental stages into the following six stages. Scribble Stage (1~3 years old) is the time when self-expression begins and draws meaninglessly. Children of this period enjoy the traces left by muscle movement rather than the purpose of drawing. In the Preschematic Stage (3-4 years old), they try to express intentionally. They express everything in a self-centered way, and they prefer emotional expressions rather than realism. In The Schematic Stage (5-6 years old), students learn the concept of objects. As interests expand, the story in the picture is enriched. Drawing Realism Stage (7~9 years old) is interested in realistic expression of form. The Pseudo-Naturalistic Stage (10-13 years old) is the time to start rational thinking. Realistic painting becomes possible, and three-dimensional spatial expression begins to appear. In the Decision Stage (13~16 years old), they discover their individuality and explore and change objects according to their intentions. It is a decision stage in which creative expression gradually disappears and many children lose interest in painting from this period [1][2]. Table 1 shows the six stages of child art development presented by Lowenfeld and a sample sketch of each stage [3].

_____

Table 1. Six stages of artistic development according to Lowenfeld [3]

Stage	Stage 1 Scribble Stage	Stage 2 Preschematic Stage	Stage 3 The Schematic Stage	Stage 4 The Drawing Realism	Stage 5 The Pseudo-Naturalistic Stage	Stage 6 The Decision Stage
Period	1-3 years old	3-4 years old	5-6 years old	7-9 years old	10-13 years old	13-16 years old
Sketch Sample						

In this study, the subjects of research are paintings aged 3 years or older who initiate conscious expression activities in an unconscious attempt. The expression of early childhood art has features that appear in common. For example, an expression of a head with a leg attached to the head, an anthropomorphic expression that humanizes non-human objects, exaggeration and reduction of self-centered forms, a clairvoyant expression of the invisible, coexistence expression that displays time and space on one screen, the use of a base line and a sky line to indicate the concept of space are representative [4].

2.2 Deep Learning and Convolutional Neural Networks

Deep Learning is a type of machine learning that finds patterns by learning a lot of data. It uses a network structure using a neural network with multiple layers. With the development of computer performance and the emergence of a lot of data, sufficient learning is possible, showing great performance in image recognition, speech recognition, intelligent robots, and natural language processing. A convolutional neural network (CNN) is a representative image recognition artificial neural network for deep learning, a convolutional layer that extracts features of an image in the intermediate layer between the input layer and the output layer, and a pooling layer that reduces the feature maps obtained from the convolutional layer. It is composed of a fully-connected layer that combines several units of each side [5-8].

2.3 Previous Studies for Hand-Drawn Image Recognition

In general, analyzing and understanding images drawn by hand or sketched using a mouse and keyboard enables applications in various fields. Image retrieval based on sketchy images is a fairly old research topic in the fields of information systems and computer science [9]. Based on a better understanding of sketch images, it can be applied to improve sketch-based image search by combining recognition results and text. Meanwhile, children's sketch drawings are known as an effective tool to better understand the child's psychological state and characteristics of development [10].

Microsoft has developed the Sketch2Tag system to recognize sketch drawings drawn by users and present results in real time. Due to the diversity and variation of hand-drawn sketches, most of the existing studies were limited to predefined classes [11]. To increase the recognition range, a collection of web-scale clipart images is used as the knowledge base of the Sketch2Tag system. Google has released 'QuickDraw', a tool that supports online sketch games. It is used as a platform for collecting large-scale sketch image data, inducing user participation and providing immediate feedback [12].

Using large scale sketch datasets such as cybertron [13] and rendered [14], including the Quickdraw dataset [12], a number of studies have been conducted on the recognition and generation of sketch images using deep learning techniques such as RNN and AutoEncoder, focusing on CNN [15-16]. In this study, we select the cybertron dataset [13] and the rendered dataset [14] to explore and evaluate the proposed model. The dataset selected here includes objects that are commonly used in everyday life because the category structure is exhaustive.

And it is recognizable only by sketch shape without additional context. In addition, categories that can be subdivided into sub-categories like 'Animal' are excluded, and sufficiently specific categories are used [13].

3. Our Method

3.1 Dataset

For children's picture recognition, two types of data sets are used: the cybertron dataset [13] and the rendered dataset [14]. The cybertron dataset consists of 80 sketch images each in 250 categories, with a total of 20,000 images. The rendered data set consists of a total of 125 categories, with 500 to 700 sketch images in each category. Among a total of 125 categories (75,481 sketch images), 47 categories were selected and the total number of data was 20,000. The category is trained with CNN and it is classified correctly. Divide the test data of the dataset by 20% and train data by 80% to train. Table 2 summarizes the characteristics of the dataset used for sketch image recognition. Sample images of each of the two data sets are shown in Figure 1 and Figure 2.

Table 2: Dataset Description

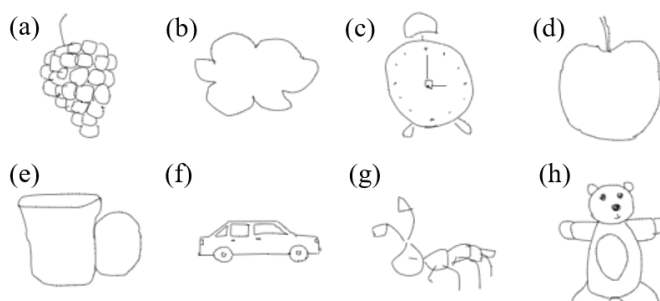| Dataset	Cybertron [13]	Rendered [14]
Number of Total Data	20,000	20,502
Number of Training Data	14,000	14,352
Number of Testing Data	6,000	6,150
Number of Categories	250	47

Figure 1. (a) ~ (h) depicts the some of drawings of the classes in sketch dataset [3] – grape, cloud, clock, apple, cup, car, ant, teddy bear respectively.
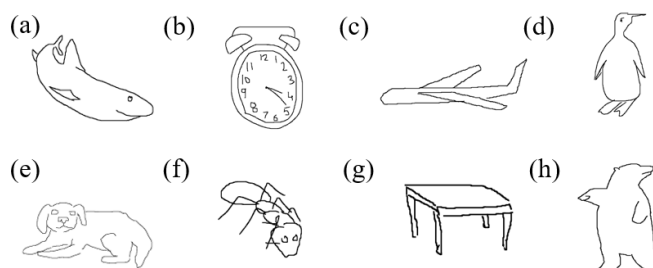

Figure 2. (a) ~ (h) depicts the some of drawings of the classes in sketch dataset [4] – shark, alarm clock, airplane, penguin, dog, ant, table, bear respectively.

3.2 Network Structure

Drawings drawn by infants are usually easily recognizable by adults. However, the process of letting the computer know what kind of picture this picture is requires processing in several stages. Before training a sketch image with

CNN (Convolutional Neural Networks), we need to make the data in an easy-to-handle data format. Therefore, the image was resized to a certain size of 64*64 and converted to a 24-bit RGB format. Therefore, one image is represented by 3 * 64 * 64 (12,288 elements in all). When constructing the CNN model, a model was created in which three layers of convolution, activation function (ReLU), and max pooling were stacked. And by placing two full-bonding layers, it finally became a 5 class. Figure 1 and Figure 2 shows a sketch image of each data set. Figure 3 is the system structure for children's picture recognition. An attractive feature of CNNs is that they serve as useful feature markers to input the output of the inner layer.
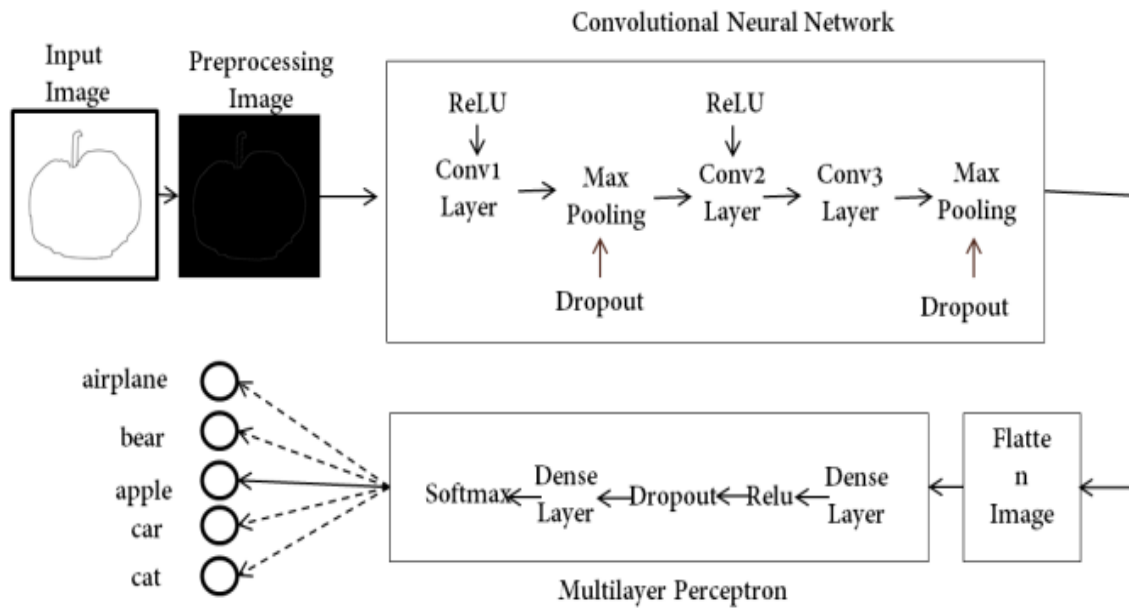
Figure 3. Structure of the system for infant figure recognition

4. Result

Table 3 shows the results of classifying about 20,000 images in each data set by CNN. Looking at the classification results, the cybertron data set showed better results than the rendered data set. The accuracy of the model's test set was about 80%. In Figure 4 and Figure 5, you can see the change in training loss and testing loss according to the epoch progress of the model. The error for the training set continues to decrease as the number of epochs increases. Finish learning before overfitting occurs.

Table 3. Performance Evaluation

Dataset	Cybertron [3]	Rendered [4]
Accuracy	0.81	0.79
Loss	0.15	0.28
Recall	0.81	0.77
Precision	0.80	0.78

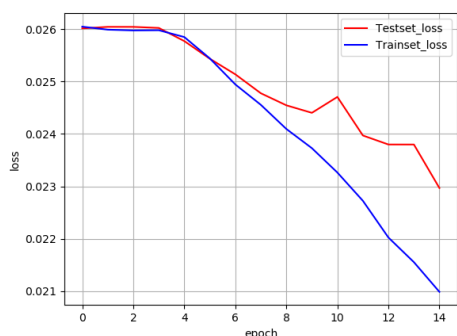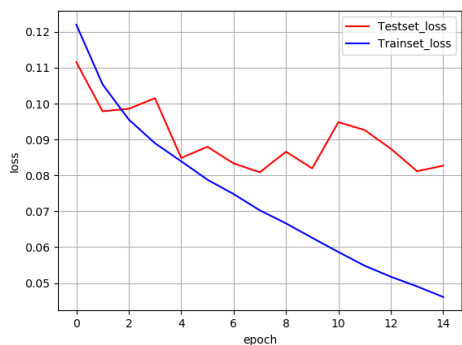Figure 4. Cybertron dataset: loss graph of learning set and test set

Figure 5. Rendered dataset: loss graph of training set and test set

5. Conclusion

This study proposed a method to improve children's drawing recognition using deep learning CNN in order to increase children's recognition of sketch drawings. The experiment accuracy was 0.81 (81%), and compared with the previous study, it was confirmed that there is some improvement effect on picture recognition. Through this, it is expected that not only the improvement of children's ability to recognize pictures, but also the measurement of learning ability and the use of children's psychological therapy through emotional recognition will be possible.

On the other hand, in order to use infant emotion recognition and psychotherapy through infant sketch figure recognition, it is necessary to define a set of diverse and detailed characteristics of sketches. For future research, we plan to conduct research on the application of image-based search and development of application tools through similarity mapping between sketch images and photographic images.

References

A. Henderson, L. L. (1978). Understanding Children's Art: Stages of Development, *Activities and Materials for Young Children. Resource Monograph* No. 22.

B. Grandstaff, L. J. (2012). *Children's Artistic Development and the Influence of Visual Culture* (Doctoral dissertation, University of Kansas).

C. The Stages of Artistic Development, *Posted on June 20, 2011 by Matt Fussell,* https://thevirtualinstructor.com/blog/the-stages-of-artistic-development

D. Coates, E., & Coates, A. (2011). The subjects and meanings of young children's drawings. In *Exploring children's creative narratives* (pp. 114-138). Routledge.

E. Yan, L. C., Yoshua, B., & Geoffrey, H. (2015). Deep learning. *nature*, 521(7553), 436-444.

F. Sainath, T. N., Mohamed, A. R., Kingsbury, B., & Ramabhadran, B. (2013, May). Deep convolutional neural networks for LVCSR. *In 2013 IEEE international conference on acoustics, speech and signal processing (pp. 8614-8618)*. IEEE.

G. O'Shea, K., & Nash, R. (2015). An introduction to convolutional neural networks. *arXiv preprint arXiv*:1511.08458.

H. Albawi, S., Mohammed, T. A., & Al-Zawi, S. (2017, August). Understanding of a convolutional neural network. *In 2017 International Conference on Engineering and Technology (ICET)* (pp. 1-6). IEEE.

I. Datta, R., Joshi, D., Li, J., & Wang, J. Z. (2008). Image retrieval: Ideas, influences, and trends of the new age. ACM Computing Surveys (Csur), 40(2), 1-60.

J. Farokhi, M., & Hashemi, M. (2011). The analysis of children's drawings: social, emotional, physical, and psychological aspects. *Procedia-Social and Behavioral Sciences*, *30*, 2219-2224.

K. Sun, Z., Wang, C., Zhang, L., & Zhang, L. (2012, October). Sketch2Tag: automatic hand-drawn sketch recognition. In *Proceedings of the 20th ACM international conference on Multimedia* (pp. 1255-1256).

L. Ha, D., & Eck, D. (2017). A neural representation of sketch drawings. *arXiv preprint arXiv:1704.03477*.

M. Eitz, M., Hays, J., & Alexa, M. (2012). How do humans sketch objects? ACM Transactions on graphics (TOG), 31(4), 1-10.

N. Sangkloy, P., Burnell, N., Ham, C., & Hays, J. (2016). The sketchy database: learning to retrieve badly drawn bunnies. ACM Transactions on Graphics (TOG), 35(4), 1-12.

O. Wang, F., Kang, L., & Li, Y. (2015). Sketch-based 3d shape retrieval using convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1875-1883).

P. Fu, L., & Kara, L. B. (2009, January). Recognizing network-like hand-drawn sketches: a convolutional neural network approach. In International Design Engineering Technical Conferences and Computers and Information in Engineering Conference (Vol. 49026, pp. 671-681).