Hand-Drawn Sketch Recognition With Double-Channel CNN

Lei ZHANG (✉ 201127001@cqcet.edu.cn)
Digital Media College of Chongqing College of Electronic Engineering

Research

Keywords: Hand-drawn sketch recognition, multi-channel, convolution neural network, deep leaning

Posted Date: May 17th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-516601/v1

Version of Record: A version of this preprint was published at EURASIP Journal on Advances in Signal Processing on August 21st, 2021. See the published version at https://doi.org/10.1186/s13634-021-00752-4.

Hand-drawn sketch recognition with double-channel CNN

Lei Zhang*

Digital Media College of Chongqing College of Electronic Engineering, Chongqing,401331, China,

201127001@cqcet.edu.cn

Abstract

In the task of hand-drawn sketch recognition, traditional deep learning methods have the insufficient of feature extraction and low recognition rate. To improve the insufficient, a novel algorithm based on double channel convolution neural network is proposed. First of all, the hand-drawn sketch is preprocessed to get a smooth sketch. And the contour extraction algorithm is adopted to get the contour of the sketch. The sketch and its contour are then used as input images of the CNN respectively. Finally, through performing feature fusion at the full connection layer, the classification results are obtained using the softmax classifier. The experimental results show that the proposed method can effectively improve the recognition rate of hand-drawn sketch.

Keywords: Hand-drawn sketch recognition, multi-channel, convolution neural network, deep leaning;

1. Introduction

With the popularity of portable touch devices, the application of hand-drawn sketching is becoming more and more diversified. More and more researchers are invested in the study of hand-drawn sketches, which includes sketch recognition [1,2,3], image retrieval based on hand-drawn sketching [4,5], 3D model retrieval based on hand-drawn sketching [6], etc.

The hand-drawn sketch recognition is still very challenging, and the reason can be attributed the following [7]. (1) The hand-drawn sketch have highly abstract and symbolic attributes. (2) Due to differences in each person's painting level and ability, the same category of objects may be very different in the shape and abstract degree. (3) The hand-drawn sketch lacks visual cues, and it has no color and texture information.

The early hand-drawn sketch recognition mainly follows the traditional image classification mode. That is, extract manual features from hand-drawn sketching, and send the features into classifier. The general manual features contains shape context features [8], scale invariant feature transformation [9] and directional gradient histogram characteristics [10] etc. However these manual features designed for natural images are not entirely suitable for abstract and sparsely hand-drawn sketch. Multi-core learning, to fuse different local features, can help improve the recognition performance, which is proved by Li Y [11]. Fisher vector (FV) is applied to recognize hand-drawn sketch, which obtained a high recognition rate [12].

In recent years, deep learning in the field of machine learning has developed rapidly. The essence of general deep learning is a nonlinear network model with multiple hidden layers. The feature expressing the original data, can be extracted from the network model to predict or classify samples by training for large-scale raw data. In the field of image recognition and computer vision, CNN has achieved the most remarkable results [13]. In addition, deep learning is widely used in pedestrian detection [14], gesture recognition [15], natural language processing [16], data mining and speech recognition. CNN can deal with the two-dimensional image directly, compared with other deep neural networks such as the depth confidence network [17] and S layer automatic coding [18]. When the two-dimensional image is converted into one, the spatial structure characteristics of the input data are lost. With the development of deep learning, some deep learning models of sketch recognition have been proposed, such as VGG [20], ResNet [19] and Alex Net [21] etc. However, these deep learning model designed mainly for color texture natural images, which does not suitable for hand-drawn sketch recognition because the hand-drawn sketching lacks of color and texture information. In the literature [22], an explicit prompt is used to require the user to draw a semantic symbol and then click the button. In the literature [23], the method of using time threshold requires users to have a clear pause after drawing a semantic symbol. In addition, special graphic symbols (such as arrows) are used for grouping [24]. The constraint conditions of these algorithms weaken the natural drawing features of the hand-drawn interface and limit the ability to express and model quickly.

* Corresponding author.

E-mail address: 201127001@cqcet.edu.cn.

The multi-channel mechanism of the CNN is used to access different views of the data (such as red, green, blue channel and stereo audio track of color images) [25]. It enables CNN to learn more abundant features and the classification effect of the model is improved, by adding input information. Therefore, in order to improve recognition rate of hand-draw sketch recognition, the contour of hand-draw sketch is also used as the input data of the CNN, then a double channel CNN is proposed in this paper.

2. Introduction

In this section, we will introduce the CNN firstly. Secondly, we present the contour extracting method of the hand-draw sketch. The proposed algorithm will be introduced at last.

2.1. Convolution neural network

CNN is an algorithm with less human intervention. The weight updating process draws on the traditional BP neural network. Error back propagation is used to update parameters automatically. Because of the lack of manual intervention, CNN can directly take images as input and automatically extract image features for identification. The features of CNN weight sharing and local perception not only reduce the number of parameters in the network, but also work in a similar way to that of animal visual nerve cells. It greatly improves the recognition accuracy and recognition efficiency of the network.

There are two typical features of CNN. The first one is the locally connection between two layers of neurons by convolution kernels rather than the fully-connection. Therefore, the convolution layer connected to the input image is a local link built for the pixel block, rather than the traditional pixel point-based full connection. Second, the weight parameter of convolution kernel is shared in each same layer. These two characteristics have greatly reduced the number of the parameters of the deep web, the complexity of the model is reduced and the training speed is accelerated. It makes CNN has a big advantage in the pixel value of processing units. The main components of CNN include the convolution layer, the pooling layer, the activation function, the full connection layer and the classifier, shown as figure 1.

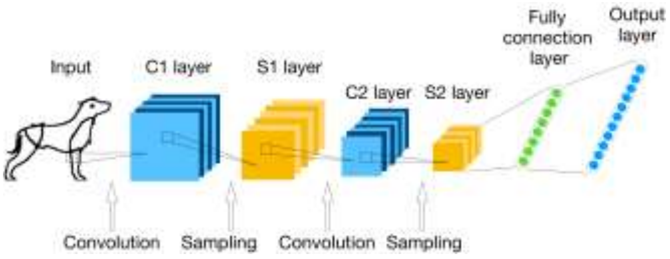

Figure 1. The structure of CNN

2.1.1. Convolution layer

The most important network layer for feature extraction in CNN is convolution layer. Convolution operation is the process of obtaining new feature graph under the action of activation function, by convolution kernel and input sample image, or upper layer output feature graph. There will be multiple feature maps at each level, which level represents a feature of the image. The operation of convolution can be represented as following.

$$x_j^l = f\left(\sum_{i \in M_j} x_j^{l-1} * k_{kj}^l + b_i^l\right) \tag{1}$$

Each level in the CNN has multiple feature maps. Suppose the jth feature map of layer l is x_j^l, where $f(\cdot)$ represents the activation function which will be described in more detail in the following chapters. M_j represents the input sample image or the set of all the input feature graphs, and k_l^{ij} represents the convolution kernel in layer l, and the convolution is expressed as $*$. After the convolution operation, we need to add the bias b after the result, then the new feature graph is formed by the activation function.

The inverse error propagation algorithm is used to update CNN weight. And the first step in the update process is to calculate the gradient at each level.

2.1.2. Down sampling and pooling layer

The down sampling layer is the process of extracting features. By lowering the dimension of feature graph for image, it is usually called pooling. In the process of down sampling, each feature graph is obtained by the dimensionality reduction of the upper layer of feature graph, which satisfies the one-to-one correspondence. Therefore, N output feature graphs of the upper convolution layer serve as input of the down sampling layer.

After dimension reduction, N output feature graphs are obtained. The following equation represents the process of down sampling and pooling.

$$x_j^l = f(\beta_j^l down(x_j^{l-1}) + b_j^l) \tag{2}$$

Where $down(\cdot)$ represents the down sampling operation. The pixel value in $n \times n$ region of the input feature graph is selected to obtain a value in the output feature graph. The dimensions of the input feature graph are reduced by n times in both horizontal and vertical directions. The final value of the pixels in the output feature graph is also related to the multiplier offset β and the additional offset b. After the activation function, the final pixel value is obtained.

2.1.3. Fully connection layer and softmax classifier

The entire connection layer is usually connected to the last layer of the pooling layer and classifier to fuse different features represented by multiple feature graphs.

Each of the neurons in the fully connection layer is connected to all the neurons in the first floor, and it has output characteristics.

The full connection layer combines all the features of the previous feature characteristics and then enters it into the softmax classifier.

After the input sample image is processed layer by layer by convolution layer and down sampling layer, a relatively complete feature set can be obtained. These features need to be classified by classifier to get the predicted value of sample image category. Then get the difference in value between predicted values and the real value. The input sample image propagates the error back through the algorithm based on gradient, so as to train the whole neural network. In general, the last layer of down sampling cannot be directly connected to the classifier, and the dimension transformation can only be used as the input of the classifier after one or two layers of full connection layer. The Softmax classifier is generally used in CNN.

Softmax classifier is suitable for multi-classification, and its prototype is a logistic regression model for binary classification. In logistic regression, assuming the sample category label is y ($y=0$ or $y=1$). There are $ m $ data samples $\{(x_1,y_1),(x_1,y_1),\ldots,(x_m,y_m)\}$ and the input characteristic $x^{(i)} \in R^{n+1}$ of these samples. The category label of the sample is 0 or 1, that is, $y^{(i)} \in \{0,1\}$, then its hypothetical function is shown as following.

$$h_g(x) = \frac{1}{1+\exp(-\theta^T x)} \tag{3}$$

Where θ is an important parameter, and θ can constitute the cost function. By adjusting parameter θ to minimize the cost function, the predicted category of the input sample can be obtained.

2.1.4. The training process

There are three main methods of CNN training, which are completely supervision, completely non-supervision, and the combination of supervision and non-supervision. The supervised learning method is used in this paper. Supervised learning is trained on the neural network in the form of a supervised signal. Supervised signal is the true value of the classification in each samples. In the learning process, the features of the input image is learned and extracted by the CNN, and the prediction value of the smaple classification is given at the output. CNN will back-propagate the difference between the predicted value and the actual value, to adjust the parameters in the network continuously. Finally, it make network for all input class will be able to make the right image of the sample.

2.2. Proposed algorithm

In this section, dand-drawn sketch contour extraction is presented firstly. Then the double-channel CNN is proposed.

The extraction process of sketch contour feature is shown in figure 2. First, the input sketch is preprocessed to obtain a smooth sketch. Then extract the contour of the sketch.

Due to the characteristics of hand-drawn sketching, it is inevitable that there will be two overlapping areas where there are redundant unclosed curvilinear segments. It is necessary to preprocess the input sketches to get a smooth contour of the sketches.

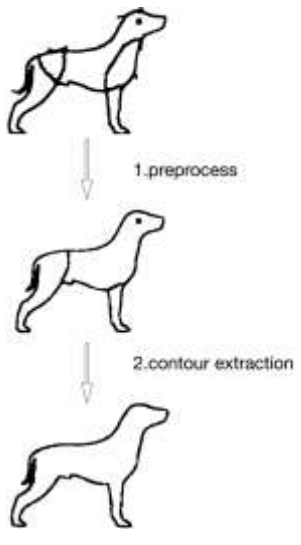

Figure 2. The process of hand-drawn sketch contour extraction

The algorithm of eliminating the unclosed curve segment can be described as following.

(a) Scan the picture according to the direction of the line. If a point is found to belong to the curve endpoint, then turn to (b). If the whole picture picture still does not have the curve endpoint, then exit. The curve endpoint can be judged when one point in a 3×3 area. If there is no other point present in the eight directions at this point, which is an isolated point, and it belongs to the curve end point. If there's only one direction has a point in eight directions, which is a little bit, then it is a curve endpoint. If there are three directions and more than three directions have a point in eight directions, then it is a curve endpoint. If there are two directions have a point in the eight directions, and the two directions are adjacent, then it is a curve endpoint, otherwise it's not.

(b) Find the endpoint of the curve, and eliminate this endpoint. Then determine whether the point adjacent to this endpoint is the curve endpoint. If it is the curve endpoint, then continue to eliminate this point and determine the next adjacent point. If it is not, then go to (a).

For the hand-drawn sketch, an adaptive tracking algorithm based on the direction of the eight-connected domain is used to extract the contour of the sketch. The original image is represented by $I(x,y)$, $C(x,y)$ represents a 2-value image of the contour. The current direction is D_i, starting from the right side and starting from the counterclockwise direction of 0,1,... 7. Select a point point_c at the left of the top line of the image as the first point. In the 3×3 area at this point, there are no other points in the upper, left and right directions. Then $d_i = 2$ is selected and look for the contour of the sketch in a counterclockwise direction. The specific algorithm is described as following.

Step1. Initialization. Set C to zero and $d_i = 2$, then the direction array DI is set to $DI =$ $\{0,1,2,3,4,5,6,7,0,1,2,3,4,5,6,7\}$.

Step2. Scan the original image I line by line from top to bottom, left to right, then the starting point of the contour is point_c can be obtained. The current point point_now is initialized to point_c.

Step3. Add the current point point_now to the binary image $C(x,y)$ of the contour. Search point_now in the order of $DI[d_i],DI[d_{i+1}],...,DI[d_{i+7}]$, to find the next adjacent boundary point point_next. If a point in one direction is found to belong to the original image I and is not equal to the initial point point_c, then this point is point_next. If the direction $DI[i]$ of the next point is found, then the new search direction d_i is the next direction in the opposite direction of $DI[i]$, that is $d_i = (\text{DI[i]} + 4 + 1)$ mod8. Then assign the value of point_next to point_now.

Step4. If the point point_now coincides with the starting point_c, then exit. Otherwise, it should be return to step3.

The effect result of this algorithm is shown in figure 2. This algorithm has good robustness. A smooth contour of the input sketch can be obtained by preprocessing the sketch.

The multi-channel mechanism of the CNN was used to access different views of the data, such as red, green, blue channel and stereo audio track of color images [26]. It enabled CNN to learn more abundant features and the classification effect of the model was improved, by adding input information. Therefore, in order to optimize the training process of hand-drawn skecth recognition, a double-channel CNN is proposed in this paper. Figure 3 shows the structure of this network. The network is composed of two relatively independent convolution network. The first input of the network is hand-drawn image, and the second is the contour of hand-drawn sketch.

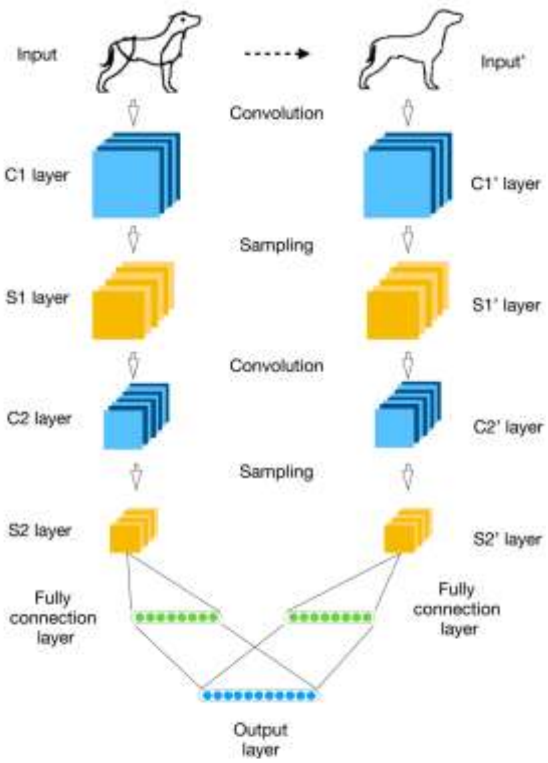

Figure 3. The structure of double-channel CNN

Each channel contains the same number of convolutional layers and parameters in a double-channel CNN, but with independent weights. After the pooling layer, the two channels are respectively connected to a full connection layer and a full connection map is performed. The two channels are connected to a fully connected hidden layer, which produces the output of a logical regression classifier. The weight of each channel has its own update. But the final error is obtained through two output layers. So, two output layers are like a deviation from each other.

3. Results and discussion

3.1. Experimental preparation

In this experiment, the configuration of computer is as follows. Windows 7, 3.60GHz, i7 processor, 32GB ddr, 1024GB hard disk. The software of experiment is Matlab 2017a.

In 2012, Eitz et al. [1] organized and collected a collection of the largest hand-sketched sketch, it contains 250 hand-drawn sketches, and each containing 80 different hand-drawn sketches. The original pixel size of the sketch is 1111×1111, as shown in figure 4. In the experiment, 4 fold cross validation was used, three for training and one for testing. The evaluation index of this experiment is the recognition rate of all test samples.

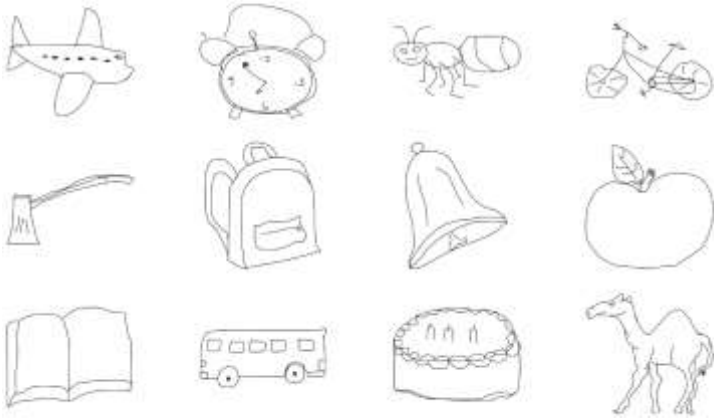

Figure 4. The sampels of hand-drawn sketches

3.2. Experimental results and analysis

Deep learning requires a large amount of training data, and the lack of training data tends to create an over-fit problem. In order to reduce the influence of overfitting, this paper makes a manual expansion of the hand-drawn sketch data set used in the experiment, and obtains a new amplified data set. Specific steps are as follows.

Step1. Dimension reduction. Reduce all the hand-painted sketch images from the original size of 1111×1111 to 256×256.

Step2. Extract the slices. From the 256×256 diagram, select five slices of the center, upper left corner, lower left corner, upper right corner and lower right corner, which size is set to 225×225. In the resulting five slices, the original dataset is made up of all 225×225 slices of pixel size in the center.

Step3. Flip horizontally. Take the five slices obtained by Step2 and flip them horizontally, and five new slices are get again. The 10 slices of each sample obtained by Step2 and Step3 constitute the amplified data set, so the data volume of the amplified data set is 10 times of the original data set.

The proposed algorithm in this paper is compared with some other popular sketch recognition methods, such as HOG-SVM [1], SIFT-Fisher [2], MKL-SVM [10], FV-SP [2] and Alex Net\cite [11]. The experimental results are shown in table 1. Compared with traditional non-deep learning methods, HOG-SVM, SIFT-Fisher, MKL-SVM and FV-SP, the recognition rate of proposed algorithm is 16.1, 9.98, 5.9 and 3.2, respectively. The results show that the depth learning method has a stronger feature and nonlinear expression than the non-depth learning mehod. Compared with the depth of the classical learning method Alex-Net, the accuracy rate is improved by 5.1. The results show that the proposed algorithm, namely double-channel CNN, can help improve the recognition rate of hand-drawn sketches.

Table 1. The comparison of recognition rate

Method	Reference	Recognition rate%
HOG-SVM	[1]	56.54
SIFT-Fisher	[2]	63.26
MKL-SVM	[10]	67.34
FV-SP	[2]	70.04
Alex-Net	[11]	68.14
Proposed	/	73.24

4. Conclusion

To improve the recognition rate of hand-drawn sketch recognition, a recognition algorithm based on double-channel convolution neural network is proposed in this paper. Firstly, perform preprocessing on the hand-drawn sketching, then extract the contour information. Secondly, the sketch and its contour are respectively used as double input channels of convolution neural network. Finally, the classification results are obtained using the softmax classifier by adopting feature fusion at the full connection layer. The experimental results show that compared with the existing mainstream methods, the proposed method of this paper achieves higher recognition rate of hand-drawn sketch.

ABBREVIATIONS

FV：Fisher vector
CNN：Convolutional Neural Network
VGG：Visual Geometry Group

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable

CONSENT FOR PUBLICATION

Not applicable

AVAILABILITY OF DATA AND MATERIAL

The labeled dataset used to support the findings of this study are available from the corresponding author upon request.

COMPETING INTERESTS

The authors declare that they have no competing interests.

FUNDING

Sichuan Provincial Department of Education Humanities and Social Sciences Key Research Base, Sichuan Hospital Management and Development Research Center 2020 Research Project (NO.SCYG2020-07)。

AUTHORS' CONTRIBUTIONS

Lei ZHANG , as the primary contributor, completed the analysis, experiments and paper writing.

ACKNOWLEDGEMENTS

Not applicable.

References

1. Eitz M, Hays J, and Alexa M, "How do humans sketch objects," *ACM Transactions on Graphics*, vol. 31, no. 4, 2012
2. Zhao P, Liu Y, Liu H, and Yao S, "A Sketch Recognition Method Based on Deep Convolutional-Recurrent Neural Network," *Journal of Computer-Aided Design & Computer Graphics*, vol. 30, no. 2, pp. 217-224, 2018
3. Seddati O, Dupont S, and Mahmoudi S, "Deepsketch: deep convolutional neural networks for sketch recognition and similarity search," *Content-Based Multimedia Indexing (CBMI), 13th International Workshop on. IEEE*, pp. 1-6, 2015
4. Liang S, and Sun Z, "Sketch retrieval and relevance feedback with biased SVM classification," *Pattern Recognition Letters*, vol. 29, no. 12, pp. 1733-1741, 2008
5. Eitz M, Hildebrand K, and Boubekeur T, "Sketch-based image retrieval: Benchmark and bag-of-features descriptors," *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 11, pp. 1624-1636, 2011

6. Li B, Lu Y, and Li C, "SHREC'14 track: Extended large scale sketch-based 3D shape retrieval," *Eurographics workshop on 3D object retrieval*, 2014
7. Yu Q, Yang Y X, and Song Y Z, "Sketch-a-net that beats humans," http://arxiv.org/abs/1501.07873v3, 2017
8. Carneiro G, and Jepson A D, "Pruning local feature correspondences using shape context," *Proceedings of the 17th International Conference on Pattern Recognition. Los Alamitos: IEEE Computer Society Press*, vol. 3, pp. 16-19, 2004
9. Lowe D G, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60 no. 2, pp. 91-110, 2004
10. Xiang Zheng, Tan Hengliang, and Ma Zhengming, "Performance comparison of improved HOG, Gabor and LBP," *Journal of Computer-Aided Design & Computer Graphics*, vol. 24, no. 6, pp. 787-792, 2012
11. Li Y, Hospedales T M, and Song Y Z, "Free-hand sketch rec- ognition by multi-kernel feature learning," *Computer Vision and Image Understanding*, no. 137, pp. 1-11, 2015
12. Schneider R G, and Tuytelaars T, "Sketch classification and classifi- cation-driven analysis using Fisher vectors," *ACM Transactions on Graphics*, vol. 33, no. 6, 2014
13. Jin L W, Zhong Z Y, and Yang Z, "Applications of Deep Learning for Handwritten Chinese Character Recognition: A Review," *Acta Automatica Sinica*, vol. 42, no. 8, pp. 1125-1141, 2016
14. John V, Mita S, and Liu Z,"Pedestrian detection in thermal images using adaptive fuzzy C-means clustering and convolutional neural networks," *Proceedings of the 14th IAPR International Conference on Machine Vision Applications on IEEE*, pp. 246-249, 2015
15. Cai J, Cai J Y, and Liao X D, "Preliminary study on hand gesture recognition based on convolutional neural network," *Computer Systems & Applications,* vol. 24, no. 4, pp. 113-117, 2015
16. Goldberg Y, "Neural network methods for natural language processing," *Synthesis Lectures on Human Language Technologies,* vol. 10, no. 1, pp. 1-309, 2017
17. Hinton G E, and Salakhutdinov R R, "Reducing the diensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504-507, 2006
18. Ranzato M A, Poultney C, and Chopra S, "Efficient learning of sparse representations with an energy-based model," *In：Proceedings of the 2007 Advances in Neural Information Processing Systems. USA: MIT Press.*, pp.1137-1144, 2007
19. He K M, Zhang X Y, and Ren S Q, "Deep residual learning for image recognition," http://arxiv.org/abs/1512.03385v1, 1th, March, 2017
20. Simonyan K, and Zisserman A, "Very deep convolutional networks for large-scale image recognition," http:// arxiv.org/abs/1409.1556v6, 1th, March, 2017
21. Krizhevsky A, Sutskever I, and Hinton G E, "ImageNet classifica- tion with deep convolutional neural networks," *Proceedings of the 25th International Conference on Neural Information Processing Systems. Cambridge: MIT Press,* pp. 1097-1105 2012
22. Kurtoglu T, and Stahovich T F, "Interpreting Schematic Sketches Using Physical Reasoning," *Proc. of AAAI Spring Symposium on Sketch Understanding. Palo Alto, USA: AAAI Press,* pp. 78-85 2002
23. Fonseca M, Pimentel C, and Jorge J, "CALI: An Online Scribble Recognizer for Calligraphic Interfaces," *Proc. of AAAI Spring Symposium on Sketch Understanding. Palo Alto, USA: AAAI Press,* pp. 51-58 2002
24. Leslie M G, Levent B K, and Thomas F S, "Combining Geometry and Domain Knowledge to Interpret Hand-drawn Diagrams,"

Computers and Graphics, vol. 29, no. 4, pp. 547-562, 2005

25. Barros P, Magg S, and Weber C, "A multichannel convolutional neural network for hand posture recognition," International Conference on Artificial Neural Networks. Springer, Cham, pp. 403-410, 2014
26. Dumoulin V, and Visin F, "A Guide to Convolution Arithmetic for Deep Learning," https://arxiv.org/pdf/1603.07285.pdf, 23th, October, 2017

Biography

Lei Zhang,,Associate Professor, Chongqing Vocational College of Electronic Engineering.

Master of Arts, Chongqing University.

Main research direction: graphic image and art design research。

Figures

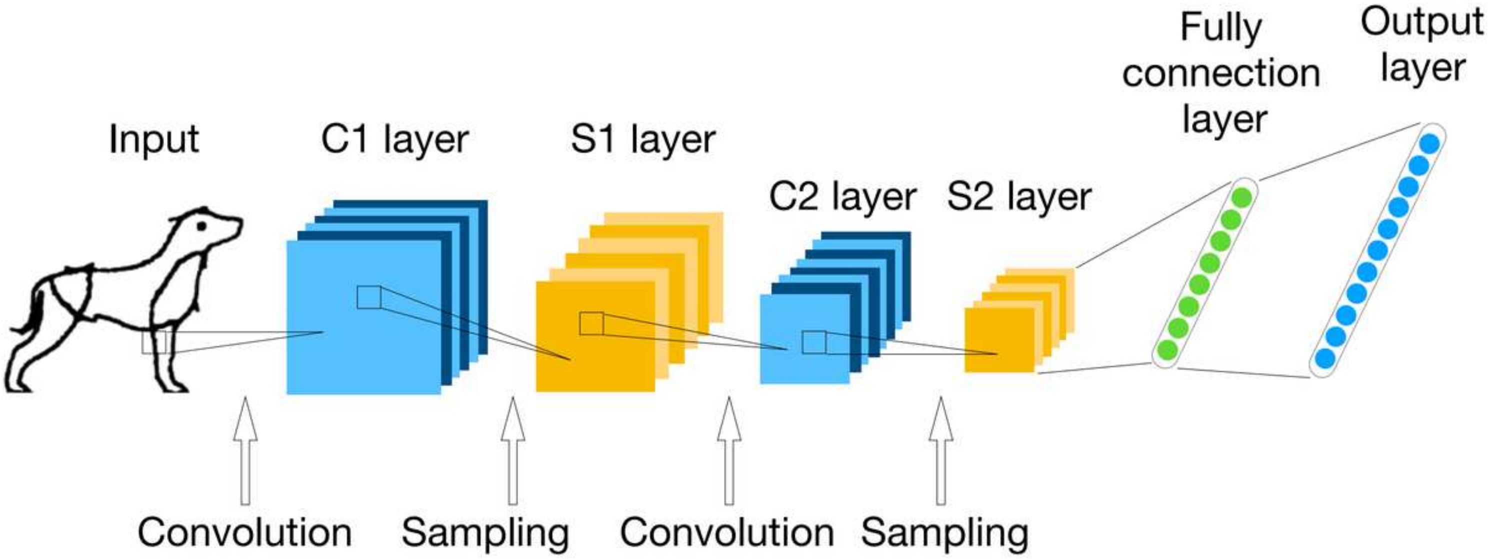

Figure 1

The structure of CNN

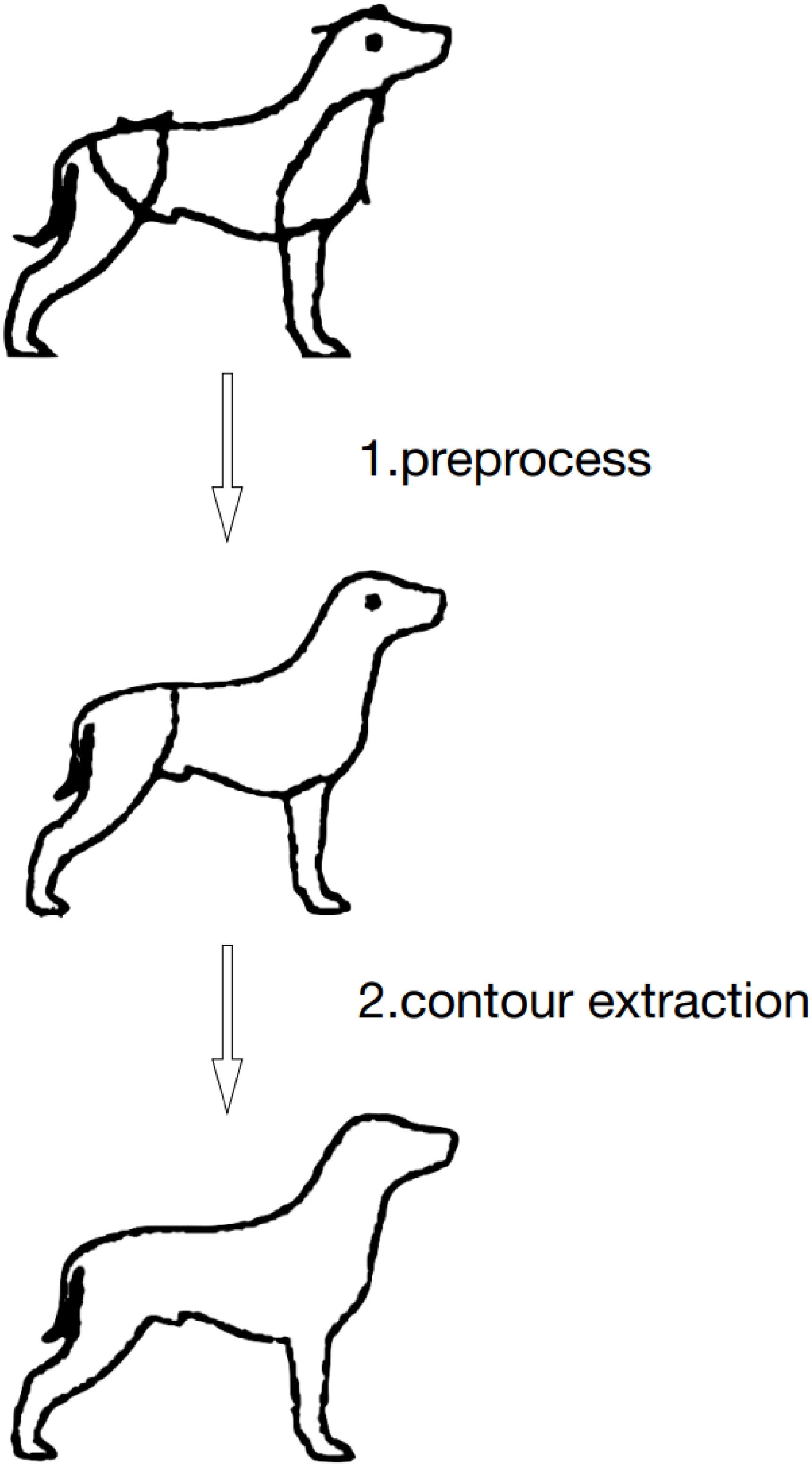

Figure 2

The process of hand-drawn sketch contour extraction

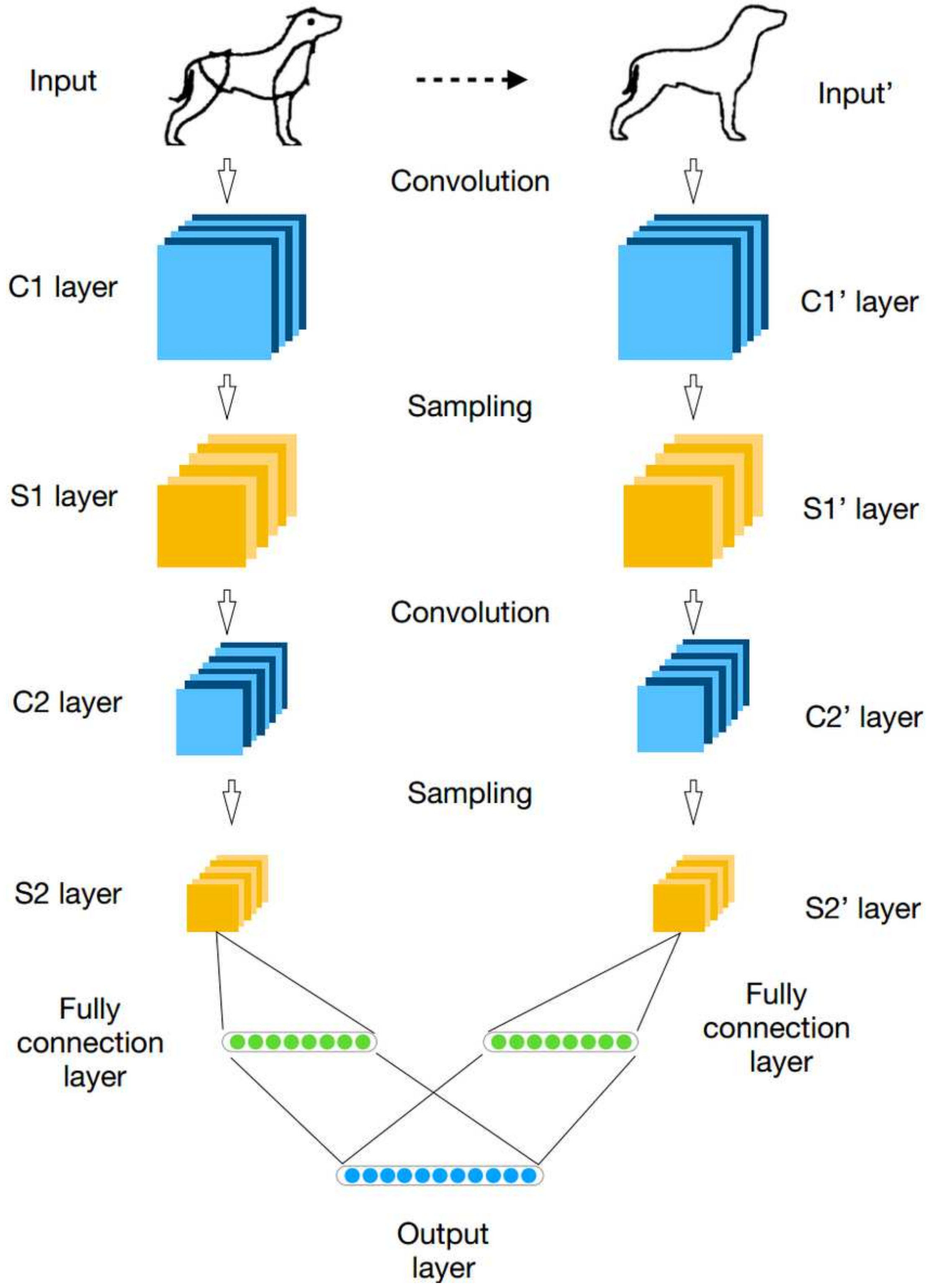

Figure 3

The structure of double-channel CNN

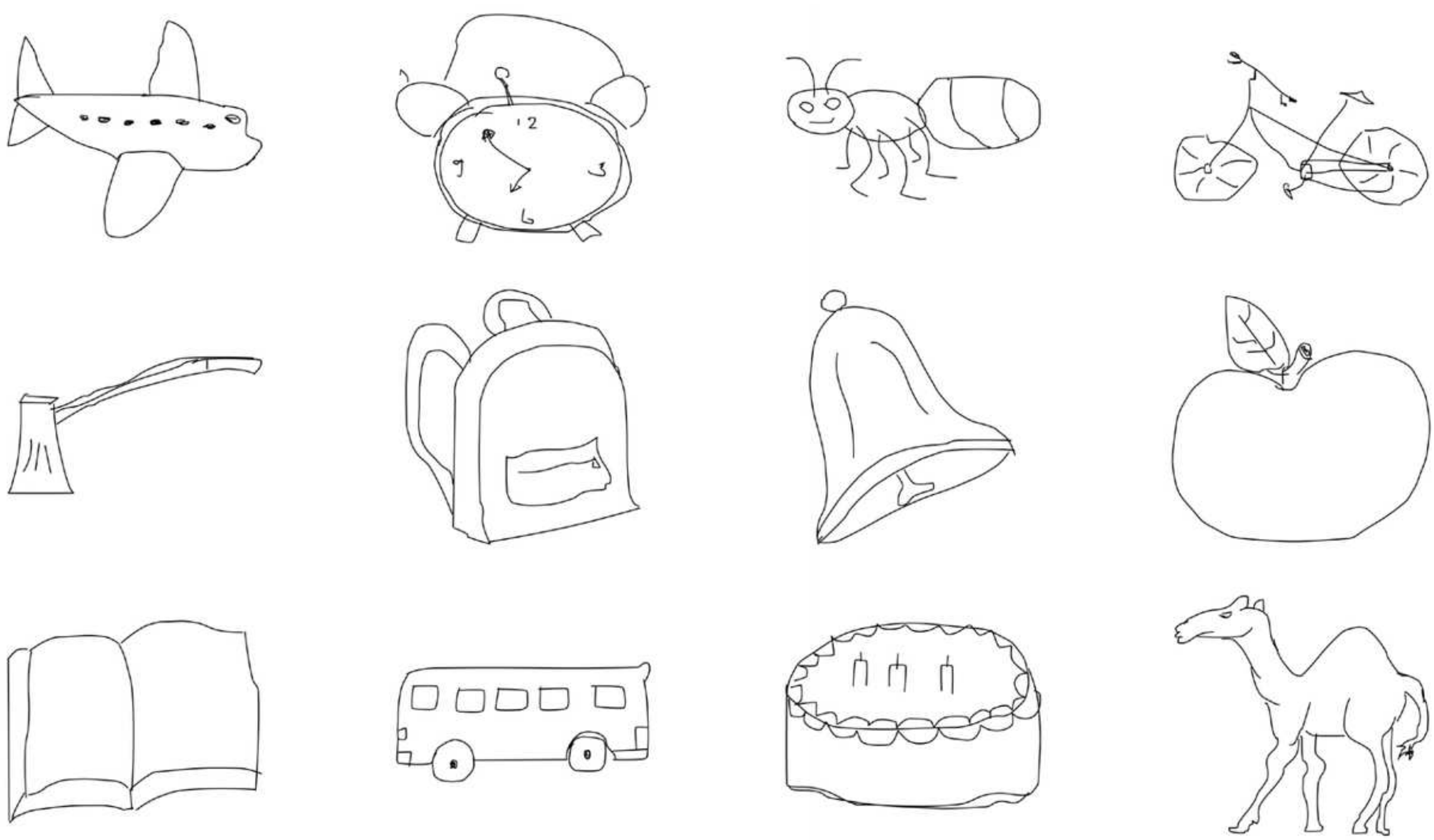

Figure 4

The sampels of hand-drawn sketches