Proceedings of the 3rd International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/4/2023368

Explore the identification of Chinese herbal medicine based on the VGG-16 model

Peiwen Zhao

Meiluyuan campus of Jiangxi University of Finance and Economics, Qingshanhu District, Nanchang City, Jiangxi Province, China

2201904211@stu.jxufe.edu.cn

Abstract. Chinese herbal medicine has a long history, is the crystallization of thousands of years of wisdom of the Chinese people, and plays a very important role in the world's medical system. Different from western medicine, Chinese herbal medicine is mainly the fruit, root system, leaves, and other parts of herbal plants, so some Chinese herbal medicine is difficult to distinguish the characteristics. In this paper, the VGG-16 model is proposed to recognize Chinese herbal medicine images automatically, which helps reduce the practical difficulty of identifying Chinese herbal medicine species. 950 images of different kinds and forms of Chinese herbal medicine were obtained through the network, and the VGG-16 model was trained by deep learning, with a final recognition accuracy of 70%. At the same time, the model was evaluated from various angles according to the main characteristics of Chinese herbal medicine, which laid a foundation for subsequent research and software development.

Keywords: Machine Learning, Chinese Herbal Medicine.

1. Introduction

Chinese herbal medicine is an important mark that distinguishes Chinese medicine from other medicine. The earliest extant monograph on Chinese medicine is Shennong Bencao Jing. For thousands of years, the Chinese people have never stopped exploring Chinese herbal medicines. Artemisia annual L. has long been widely used in the folk treatment of malaria in China. In the 1970s, Chinese traditional medicine scientist Tu Youyou and her team isolated a new sesquiterpenoid containing peroxide group from the main chemical components of Artemisia annua. Arteannuim [1], and proved that it is an effective component of antimalarial work for the world made a great contribution. In the outbreak of SARS in 2003, TCM and Chinese herbal medicine also made outstanding contributions to the world's anti-SARS campaign.

With the continuous development of modern medicine, more and more people have realized the importance of Chinese herbal medicine. In China, the research of Chinese herbal medicine is not limited to the pharmacological research of Chinese herbal medicine and the development of new Chinese medicine. Chinese herbal medicine has its unique role in various fields in China. Will have mouldproof bacteriostatic action of Chinese herbal medicine made of plastic wrap can effectively inhibit the deterioration of fresh fruit, to extract the effective components of Chinese herbal medicine used in feed additive can be an as effective alternative to forage antibiotics, meat processing used in Chinese herbal medicine made of natural antioxidant can effectively guarantee the product quality and

DOI: 10.54254/2755-2721/4/2023368

safety [2]. Chinese herbal medicine has attracted the attention of scholars worldwide because of its unique historical and natural nature.

According to statistics, at present, China has a total of more than 400 kinds of commonly used Chinese herbal medicine [3], among which Guangxi and Gansu provinces are the main origins of Chinese herbal medicine. In 2009, the output of Traditional Chinese medicine in Gansu province reached 503,600 tons, with an output value of 3.03 billion YUAN [4]. The huge economic prospect has also spawned more and more large traditional Chinese medicine enterprises. In addition to medicine farmers breeding Chinese herbal medicine, wild Chinese herbal medicine is due to its high price to attract a large number of people. However, the imbalance between the supply and demand of wild Chinese herbal medicine in the market often leads to the rising market price of wild Chinese herbal medicine. More and more people join the team of mining wild Chinese herbal medicine, resulting in the circulation of a large number of poorly made herbs in the market at the same time.

Currently, some software has been used to identify Chinese herbal medicine in the market. The recognition method is mainly imaging recognition, and the correct information can be analyzed by analyzing the shape of herbal medicine in the image. Among them, the electronic nose, namely the odor scanner, is unique in the recognition of Chinese herbal medicine. It integrates a variety of sensors and uses an artificial neural network to simulate the olfactory nerves of living creatures [5].

However, at present, identification tools including electronic nose are difficult to achieve sufficient accuracy in practical application, and electronic nose is also difficult to accurately apply in the field of Chinese herbal medicine identification due to the difficulty in selecting sensors [5]. The problem is that most of the existing identification tools with high accuracy cannot deal with the identification of herbal medicines alone, and the tools developed based on the identification of herbal medicines are difficult to guarantee accuracy. Due to the large subdivision of Chinese herbs and the variety of forms brought about by the different degrees of grinding, we need a recognition tool dedicated to this domain, and the ability to analyze and identify Chinese herbs by training specific models through large data sets. To verify the feasibility and accuracy of the Chinese herbal medicine identification tool designed based on the VGG-16 model, this experiment adopts a quantitative research method to carry out a targeted model design and accuracy experiment according to different forms of five common Chinese herbal medicine. The identification accuracy is the final judgment basis of this experiment, and the feasibility of the model in the direction of Chinese herbal medicine identification is studied and analyzed, and the data support for the subsequent development of practical tools is provided.

2. Literature Review

2.1. Chinese Herbal Medicine

Chinese herbal medicine has a wide range of sources, high curative effect, is easy to use, easy to draw materials, etc. It mainly comes from 39 subjects of angiosperms, and each herb has different pharmacological effects and morphological characteristics. And Chinese herbal medicine has been an important medicine against plague since ancient times, but its safety has always been difficult to be guaranteed. Due to the particularity of traditional Chinese herbal medicine, such as the combination of medicinal effects, many medicines achieve ideal effects through the combination of various herbs, which means that there are large side effects. Different from western medicine, the form of Chinese herbal medicine is mainly pasted, pill, powder, and soup. In addition, the existing patent law for traditional Chinese herbal medicine is not perfect, and many widely spread Chinese herbal medicine can not be protected by patents. Therefore, a large number of fake and shoddy products flow into the market, which greatly affects the spread of Chinese herbal medicine [6]. At the same time, too many fake and shoddy products have seriously affected the safety of Chinese herbal medicine. To improve the fluidity of all kinds of data on Chinese herbal medicine, there are relatively complete databases in the world to update the latest information on Chinese herbal medicine products in time. Besides, There are already electronic noses on the market that identify Chinese herbs that mimic the way human

DOI: 10.54254/2755-2721/4/2023368

organs perceive their scent [7]. However, the practical application of the electronic nose is not enough to illustrate its efficiency in drug quality control [8].

2.2. Images identification based on machine learning

In recent years, intelligent recognition function based on machine learning has been widely concerned and has a relatively mature application. Given the reduced accuracy of image recognition caused by more human intervention in the early stage, researchers have proposed a variety of new Network structure models, such as Deep Belief Network, Convolutional Neural Network, and so on. Besides, face recognition, as a prominent field, has been applied with high frequency in the field of public security, and micro-expression judgment has greatly improved the practical application range. Although the accuracy of face recognition has been greatly improved in the past, the accuracy of the algorithm is still not guaranteed without good pictures [9]. Compared with face recognition software, image recognition is carried out on the premise of taking HD images on the spot, and the analysis accuracy of is higher. To obtain higher recognition accuracy, this experiment uses the existing high-precision image recognition model to realize the recognition of Chinese herbal medicine.

3. Methodology

This experiment adopts a quantitative research method. The data sets of high definition and complete types of Chinese herbal medicine were obtained from the latest Chinese herbal medicine magazines and authoritative image websites such as VGG.COM, and these data sets were sorted according to the morphological aspects of herbal medicine, etc. Then the vGG-16 Chinese Herbal medicine intelligent recognition model was used to analyze and learn the training set. Finally, the accuracy of the model for Chinese herbal medicine image recognition was judged.

This experiment aims to explore the accuracy of the image recognition function based on machine learning in the image recognition of Chinese herbal medicine. For the deep learning model, VGG-16, named after the Visual Geometry Group, is selected. The core of the model is five convolution operations on the data set. Max-pooling space dimension reduction is carried out between every two groups, and continuous 3 x3 convolutions are used for multiple times in the same group. The number of convolution kernels increases in shallow groups from 64 to 512. The number of convolution kernels in the same group is the same as that in the most serious groups. After convolution you have two completely connected layers, and then you have the classification layer. VGG has a simple structure and adopts the combination of a small convolution kernel and multi-convolution sub-layer, all of which adopt a 2X2 pooled kernel. At the same time, increasing the number of channels layer by layer can effectively extract more information.

Figure 1. Image samples of the Chinese Herbal Medicine dataset.

In terms of data set selection, lily, Codonopsis, Lycium barbarum, Sophorae japonica, Honeysuckle, and five kinds of common Chinese herbal medicine honeysuckle were mainly used in this experiment. Images with high definition and obvious characteristics of Chinese herbal medicine were screened from VGG.COM, and a total of 950 clear images were included. To explore the accuracy of the experiment, samples of five different forms, including intact root system, intact petals, single petals,

DOI: 10.54254/2755-2721/4/2023368

and chopped herbs, were selected. Finally, these data sets are divided into training sets and validation sets, and the data list is generated out of order. After the data is prepared, configure the VGG-16 model and train the model using the training set.

Through quantitative research, the accuracy of experimental results can be guaranteed by selecting samples of commonly used Chinese medicinal materials and providing a sufficient sample size during model training when the total amount of Chinese medicinal materials is large.

4. Result

Table 1. Results of model prediction.

Name	Main Form	Prediction
Sophorae japonica	Leaf	100%
Honeysuckle	Leaf	50%
Lily	Petal	50%
Codonopsis pilosula	Root	100%
Lycium barbarum	Seed	100%

The total number of samples in this experiment was 950, and the sample pictures included lily, Codonopsis, Lycium barbarum, Sophora japonica, and Honeysuckle. In the model evaluation, these images were divided into four patterns: root, seed, Leaf, and petal. It is worth mentioning that the main form of lily is a petal, the main form of Codonopsis is the root, and the main form of Lycium barbarum is the seed, the main form of Sophora japonica is the petal, and the main form of Honeysuckle is the leaf.

From the perspective of herbal species, the accuracy rate of the model recognition is 70%, and the accuracy rates of Sophora japonica, Codonopsis, and Lycium barbarum are 100% similarly. The accuracy rate of lily is 50%, while honeysuckle is difficult to be accurately identified by the model, its accuracy rate is 0. From the perspective of herbal morphology, the model could accurately identify petals and roots, while the leaf had a high degree of confusion in the model and was easily mistaken as another petal.

In addition, we also found a problem worth thinking about. In this experiment, the model recognized sophora japonica 50% of the time, which means that except for the correct samples, all the wrong identification results were recognized as sophora japonica by the model.

5. Discussion

The results showed that the image characteristics of Chinese herbal medicine had a great influence on the identification accuracy, and the number of herbs per unit was Root>Seed>Leaf>Petal, with the accuracy decreasing in descending order. The reason for this result is that the individual size of these four forms decreases successively. The smaller the individual is, the more herbaceous plants there are in each unit, so the corresponding pixel value is lower, and the model recognition is more difficult. In addition, the VGG-16 model was evaluated by comparing the images provided in the training set. It is not difficult to know that in the process of identifying the species of Chinese herbal medicine in the image, the model needs to separate and analyze the individual Chinese herbal medicine in the image. Therefore, under the premise of the same image pixels, The identification accuracy of the VGG-16 model for Chinese herbal medicine mainly depends on the number of pixels occupied by each Chinese herbal medicine individual, because the higher the pixel, the clearer the image. Planting herbs, on the other hand, the integrity of the individual form also to some extent affects the efficiency of model identification, when the overlapping between each part of the image is large, the model is difficult to accurately obtain the image of Chinese herbal medicine correct morphological characteristics of the individual, and the recognition of individual sample will greatly increase the difficulty, further analysis can know, The smaller the individual medicinal herbs are, the lower the individual pixels are,

DOI: 10.54254/2755-2721/4/2023368

the more chaotic the position of individual herbs in the image will be, and the more overlap between them will be, ultimately leading to the difficulty of accurately capturing their overall shape in the model, thus reducing the accuracy of recognition. In contrast, the VGG-16 model can accurately identify the herb category if the individual in the image is large and there is not much overlap.

Since the number of samples identified as sophora japonica reached 50% of the total number, we can understand that some characteristics of Sophora japonica are easier to be identified by the model. Further analysis, we found the difference between the sophora japonica and the other four kinds of Chinese herbal medicine is that the morphology of sophora japonica is more rich. Besides the main shape of it is the petal, also is similar to the seed, leaf, etc, which means that the model incorrectly identified other types as Sophoricae japonicae in the process of recognition and analysis, therefore in this experiment, Sophora japonica has high accuracy and high recognition probability. This result just shows that the rich form of Chinese herbal medicine is the biggest difficulty in designing and developing high-accuracy Chinese herbal medicine recognition software, which is consistent with the analysis results in the early stage of the experiment.

6. Conclusion

The main purpose of this study was to explore the accuracy of the VGG-16 model in the identification of Chinese herbal medicine. Based on the quantitative study of the morphological characteristics of Chinese herbal medicine, the factors affecting the accuracy of model recognition were obtained. This result indicated that the VGG-16 model could effectively identify Chinese herbal medicines in large individuals, but it was less efficient to identify small and overlapping parts of individuals. Despite this, the model still has some reference significance in the field of Chinese herbal medicine recognition. The VGG-16 model is effective in identifying individuals in high-resolution images, but some herbs are small enough that the resolution assigned to each individual is limited. Therefore, if the shape of each individual can be captured or each individual can be arranged in an orderly way as far as possible in practical application, the model can still achieve excellent results.

This experiment confirmed the operability of the VGG-16 model in the identification of Chinese herbal medicine, but it did not reach the ideal accuracy. Based on some existing conclusions, the optimization of the model data set is a focus of the follow-up research.

In this experiment, the feasibility and accuracy of the VGG-16 model in the identification of Chinese herbal medicines were verified and evaluated by high-definition images of five common Chinese herbal medicines, which confirmed the operability of the model in this field, and provided more comprehensive data support for the subsequent development of practical tools.

References

[1] Liu J, Mi M, Fan J, Tu Y, Wu Z, Wu Y and Chou W 1979 Structure and reaction of Arteannuin. *Acta Chimica Sinica.* 02:129-143.

[2] Mao Q, Song X and Luo Z 1999 *Food Science.* 20(5):54-56. p55

[3] Huang S and Ye A 2016 *Rural Economy and Science-technology* 19:21-24. p21

[4] Han D, Wang H, Wang Y and Liu G 2011 Present situation of Chinese herbal medicine production and sustainable development strategies in Gansu province *Chinese journal of agricultural resources and regional planning* 06:56-60. p56

[5] Wang J, Hu G, Yu Y and Zhou Y 2004 Research and application of electronic nose and electronic tongue in the food inspection *Transactions of the Chinese Society of Agricultural Engineering* 29(14):292-295. p294

[6] Yang S, Wu N, Yuan X, Liu Y, Zhong R and Wu C 2014 Current situation and thinking on "odor and taste" identification of traditional Chinese medicine *World Sci. Technol. Mod. Tradit. Chin. Med.* 16(9):1876–1879. pp1876-1877

[7] Gebicki J. 2016 Application of electrochemical sensors and sensor matrixes for measurement of odorous chemical compounds *TrAC Trends Anal. Chem.* 77:1–13.

DOI: 10.54254/2755-2721/4/2023368

[8] Huang M 2016 Current situation and development trend of traditional Chinese Medicine *Lishizhen Med.Mater.Med.Res.* 8:1956–1960.
[9] O'Toole AJ, An X, Dunlop J, Natu V and Phillips P.J 2012 Comparing face recognition algorithms to humans on challenging tasks *ACM Trans. Appl. Percept.* 9:1-13. pp1-4.