Chinese Herbal Recognition Databases Using Human-In-The-Loop Feedback

Nan Wu
Yunnan Medical Health College, Kunming Yunnan, China and School of Medicine, Yunnan University of Business Management, Kunming Yunnan, China
wunan2885@gmail.com

Yujun Zhou
School of Medicine, Yunnan University of Business Management, Kunming Yunnan, China
1062726102@qq.com

Hao Xu
School of Medicine, Yunnan University of Business Management, Kunming Yunnan, China
2846234659@qq.com

Xinjiao Wu
East China Normal University, Shanghai, China
wuxingjiao2885@gmail.com

ABSTRACT

Traditional Chinese medicine identification plays an important role in the development of traditional Chinese medicine. Traditional Chinese medicine identification mostly relies on researchers' experience, so traditional Chinese medicine identification is still challenging. Using the computer identification of traditional Chinese medicine seems an effective method, but no dataset can train models. The lack of a dataset is the challenge of traditional Chinese medicine identification by use computers. This paper proposes a method for constructing a Chinese medicine dataset based on human-in-the-loop. This method uses a manual intervention labeling method to realize a labeling mode that saves labour resources. First, we use a web crawler to collect data from the Internet, then use a pre-model to remove some irrelevant data, next, we iterative data annotation based on the classification confidence, finally, we will obtain a dataset named CH42 that annotation by human-computer collaboration. Besides, we designed a backbone network for explicitly modeling interdependencies between channels. The CH42 contains 42 types of Chinese medicine data, a total of 6,112 pictures, the model automatically labeled about 64% of the data. We sampled 6 sets of data and found 6 mislabeled data from 1458 pictures. The model labeling accuracy rate is about 98.6%.

CCS CONCEPTS

• **Human-centered computing**; • **Human computer interaction (HCI)**; • **HCI theory, concepts and models**;

KEYWORDS

Traditional Chinese medicine identification, Human-in-the-loop, Dataset, Deep learning

CSAE 2021, October 19–21, 2021, Sanya, China

ACM ISBN 978-1-4503-8985-3/21/10…$15.00
https://doi.org/10.1145/3487075.3487114

ACM Reference Format:
Nan Wu, Yujun Zhou, Hao Xu, and Xinjiao Wu. 2021. Chinese Herbal Recognition Databases Using Human-In-The-Loop Feedback. In *The 5th International Conference on Computer Science and Application Engineering (CSAE 2021), October 19–21, 2021, Sanya, China.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3487075.3487114

1 INTRODUCTION

As an indispensable part of traditional Chinese medicine treatment, traditional Chinese medicine has played an important role in the development of medicine. Among them, the identification and classification of traditional Chinese medicine is a relatively challenging task. At present, the identification and classification of traditional Chinese medicine are mostly based on the experience of researchers, similar to the family inheritance model and the systematic teaching of professional universities, which makes the research and development of traditional Chinese medicine subject to domain knowledge limit. It will provide new impetus for Chinese medicine development if can provide a Chinese medicine identification model.

Image classification is the key technology to solve the above problems [1]. Image classification aims to construct a depth model by using technologies such as convolutional neural networks and use the model to determine the category of a given input image. However, using image classification technology is a challenge because no dataset can train Chinese medicine classification. Besides, the current labeling cost is very expensive, and the labeling of traditional Chinese medicine identification datasets can only be performed by professionals. Therefore, whether a more convenient labeling method can construct a traditional Chinese medicine identification dataset is an important task. The core issue that this paper focuses on is: *Can we used a more convenient way to build a Chinese medicine identification dataset?*

The human-in-the-loop approach uses machine learning methods to train a model from many training data or a small number of samples and uses the model to predict new data. When the prediction confidence is too low, people take the initiative to intervene and make judgments [2]. Due to the large similarities in the data of traditional Chinese medicine, multiple iteration labeling methods can be adopted, and the uncertain samples are selected for further labeling each time.

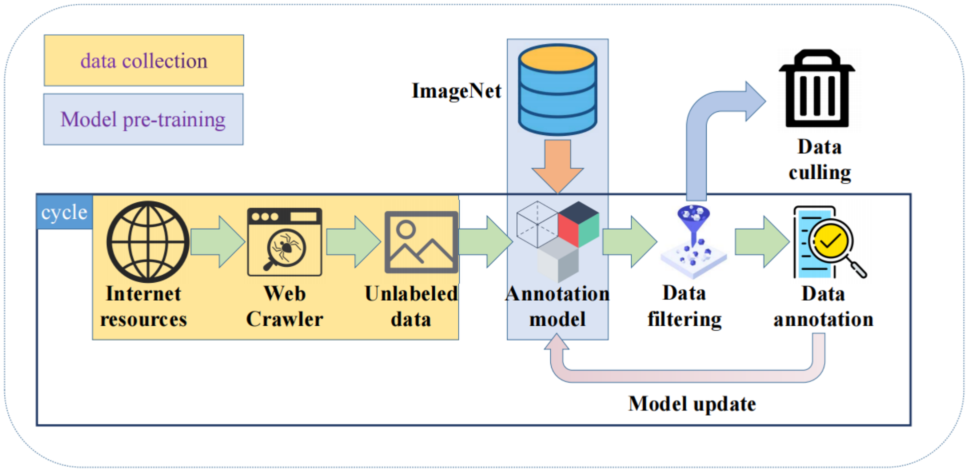

Figure 1: The Framework.

This paper designs a way of labeling based on the human-in-the-loop, (1) Determine the label category with experts in the field of pharmacy; (2) Use web crawlers to grab pictures from the Internet through keyword crawling; (3) Use ImageNet The trained model screens out irrelevant samples; (4) The pharmacy expert selects many data in the dataset for labeling; (5) Performs the first round of iteration, fine-tune the model, and screens out uncertain samples and submits them to experts for labeling; (6) Perform the first round of iterations to fine-tune the model. In the second round of iteration, uncertain samples are selected for labeling; (7) Obtain the labeled dataset. The framework is shown in Figure 1

Contributions to this paper are as follows:

- We propose a solution of constructing a dataset based on human-in-the-loop, which is to use the fewer label data to construct a new dataset. This solution can be used not only to construct a Chinese medicine dataset but also to construct other new domain datasets.
- We constructed a new standard Chinese herbal medicine dataset, which not only supports the identification of Chinese herbal medicines but also supports the research of machine learning methods.
- As far as we know, we are the first to propose using a human-in-the-loop in the construction of a Chinese medicine dataset.
- We have proved the rationality and superiority of the framework proposed in this paper through experiments.

2 RELATED WORK

2.1 Human-in-the-Loop

The purpose of the human-in-the-loop is how to provide training data for machine learning. Simultaneously, the human-in-the-loop also focuses on how to use human power to complete certain tasks that are difficult for computers to complete. At present, the human-in-the-loop has achieved milestone effects in various fields. Li *et al.* [3] established a simulator to simulate human-machine dialogue in the natural language processing task. The robot can improve the question and answerability based on the teacher's feedback finally used the Turk robot to verify the method. Plummer *et al.* [4] introduced an attribute-based interactive image search method that uses the human-in-the-loop to optimize image search results in computer vision tasks iteratively. Böhme *et al.* [5] innovatively introduced a semi-automatic BUG repair framework for human intervention, which uses a human-in-the-loop method to train an automatic BUG discovery model to liberate the workforce more.

Renzaihuan has much excellent work in dataset construction, Yu *et al.* [6] used the human-in-the-loop method to iteratively label a dataset containing 10 scenes and 20 object categories. This dataset contains approximately one million labeled images. Guoliang Li [7] proposed a human-machine hybrid data integration framework to deal with inconsistent data sources in crowdsourcing tasks and developed a crowd-driven database system on the crowdsourcing platform.

In order to cope with the challenge of relying on large amounts of manually labeled data for deep learning, Zhang *et al.* [8] proposed a framework that combines the advantages of regular expressions and deep learning. Many tasks have almost no annotated data and always need to be created from scratch. To cope with this challenge, Klie *et al.* [9] proposed a novel method of domain-agnostic circle human annotation. Yue *et al.* [10] proposed a framework called "Interventional Learning with Few Students (IFSL)" to solve an overlooked flaw in the recent Few-Shot Learning (FSL) method. Wan *et al.* [11] proposed a Human-In-the-Loop Low-shot (HILL) learning algorithm to dynamically reject uncertain predictions and label them to add a set of novel categories.

2.2 CNN Applied on Tasks Like Herbal Recognition

At present, there is little research literature on Chinese herbal medicine identification. Xu *et al* [12]. proposed a multi-attention pyramid network for Chinese herbal medicine identification, and at the same time proposed the dataset named CNH98, but this dataset is not open source. In addition, their datasets are fully supervised and annotated, instead of using this human-machine hybrid mode for annotation. In addition, We have not searched for work on the construction of Chinese medicine datasets, and many similar works deal with plant identification [13], leaf identification [14], and flower identification. As far as we know, there is no work on dataset construction using the human-in-the-loop approach [15].

3 METHODS

The Chinese herbal medicine recognition framework based on human-in-the-loop proposed by us is shown in the figure. Our framework can be divided into four parts. The first part is the data collection module, and the second module is the model pre-training module. The selected model will become the core of the entire task. We use a SeNet-18 as the selected model. The SeNet adaptively recalibrates channel-wise feature responses by explicitly modeling interdependencies between channels. The third part is the data filtering and labeling module, and the last part is the model update module.

3.1 Data Collection

The Internet is full of various digital resources, but these resources need to adopt an effective automatic collection method. Following the most effective data collection method at present, we choose to use web crawlers for data collection. Before starting the data crawling, we first consulted a number of pharmacy researchers and then used voting to determine the 42 most commonly used Chinese medicines as the dataset elements. Based on the python 3 platform, we use selenium to crawl data from Google browser, and we crawl data from Bing. We define the dataset to be crawled as N_{CH42}.

$$N_{CH42} = \bigcup_{i=0}^{42} \{x_b\} \tag{1}$$

$x\in${Sanchi, Tall Gaxtraodia Tuber, Lucid Ganoderma, Saffron, Chinese Angelica, Indian Buead Tuckahoe, Ginseng, Chinese Caterpillar Fungus, Chinese wolfberry root-bark, Coix lacryma-jobi, ginkgo seed, Wild Mint Herb, Noble Dendrobium Stem Herb, Sickle Senna Seed, Safflower, Indigowoad Leaf, incised notopterygium rhizome and root, rhubarb root and rhizome, Indigowoad Root, Root of Indigowoad, Weeping Forsythia Capsule, Common Anemarrhena Rhizome, reed rhizome, Cablin Potchouli Herb, angelica dahurica, Opium Poppy, Platycodon Grandiflorum, Cherokee Rose Fruit, Crataegus Pinnatifida, Fresh Ginger Sargentgloryvine Stem, Nutgrass Galingale, Black Sesame, Walnut Meat, White Paeony Root, Common Bletilla Tuber, Taraxacum, Lobed Kudzuvine Root, Tuber Fleeceflower Root, liquorice root Tangerine Peel, Wild Honeysuckle Flower, Mongolian Dandelion Herb}.

The number of images we crawled was 20,360.

3.2 Model Pre-Training

The recognition model is a key part of the whole framework. We select the SeNet-18 model with channel feature extraction as our recognition model. We define M as the $n\times n$ confusion matrix with n categories. We evaluate our model using accuracy, precision, recall, and F_1 as metrics (Eq. 2),

$$\begin{aligned} Acc &= \frac{\sum_i M_{ii}}{\sum_{ij} M_{ij}} \quad P_i = \frac{M_{ii}}{\sum_j M_{ij}} \\ R_i &= \frac{M_{ii}}{\sum_j M_{ij}} \quad F_1 = \frac{2\times \sum_{i=1}^{n} P_i + \sum_{i=1}^{n} R_i}{n(\sum_{i=1}^{n} P_i + \sum_{i=1}^{n} R_i)} \end{aligned} \tag{2}$$

In order to select a recognition model with better performance, we select 30 samples from each category to form a dataset. Split the training set and test set according to 8: 2 to evaluate several common models at present, and each model is pre-trained using ImageNet [16] in advance. The results are shown in Table 1

It can be seen from Table 1 that the SeNet structure has good performance. The reason for the analysis is that the characteristics of SeNet modeling channels can enhance the effective information of the channel while suppressing the invalid information of the channel. So we finally chose SeNet-18 as our recognition model.

As shown in Figure 2, the main part of SeNet-18 is a common ResNet18 model, and the SeNet structure is added to the jumper module. In this way, it can be ensured that the feature correlation between channels can be effectively modelled during data fine-tuning.

As shown in Figure 2, for a characteristic channel CH1 with a length of C, a width of W, and a height of H, after compression, expansion, and laying operations, the channel-interdependencies are automatically obtained. We can use the channel interdependencies can enhance useful features and suppress less useful ones.

To ensure the feature extraction capability of the model, we first use ImagNet to pre-train the model. Since the data crawled from the web contains a lot of noise data, this part of the data is scattered in various parts of the dataset. These noise data are roughly divided into four categories, humans, plants, medicine boxes, and food. The model can filter these data. We first searched for 100 human, plant, medicine boxes, and food data to fine-tune the model. Then use the model to identify all crawled datasets, and take the model's softmax output as the confidence level. Remove all data that the class belongs to humans, plants, medicine boxes, and food with a greater confidence level than 0.7. The number of the filtered images by using pre-training models show as Figure 3

3.3 Data Filtering and Labeling

After screening by the pre-training model, 7327 pieces of data were deleted. The strategy of deleting high-confidence classification data is to ensure that the deleted data are all wrong. Next, we will filter out 30 images from each category to train a model pre-trained by ImageNet. Then screen out samples with a confidence level of less than 0.3 for artificial labeling. Then the artificially labeled samples are sent to the model for training, and the data is sent to the model for iteration. We select samples with confidence less than 0.3.

3.4 Dataset

After 4 rounds of data iteration, our model has no sample output with confidence less than 0.3. At this time, we have marked the completed dataset. We show some samples in Figure 4, and at the same time, we show the final number of labeled data in Figure 5

Table 1: Comparison with Recently Proposed Methods

Method	*Depth*	*Params(M)*	*top-1(%)*	*P*	*R*	F_1
Network in Network [17]	-	2.3	81.3	0.786	0.767	0.776
VGG-11[18]	11	129.1	84.2	0.806	0.819	0.812
ResNet-18[19]	18	11.7	91.5	0.907	0.893	0.899
WRN-18-1.5[20]	18	25.2	92.5	0.915	0.921	0.918
CBAM-ResNet-18[21]	18	11.3	91.8	0.913	0.930	0.921
FPN-ResNet-18[22]	18	13.3	92.1	0.909	**0.946**	0.927
VGG-19[19]	19	140.0	85.9	0.846	0.835	0.840
GoogleNet[23]	22	7.0	91.2	0.880	0.829	0.854
SeNet-18[16]	18	11.8	**93.7**	**0.926**	0.945	**0.935**

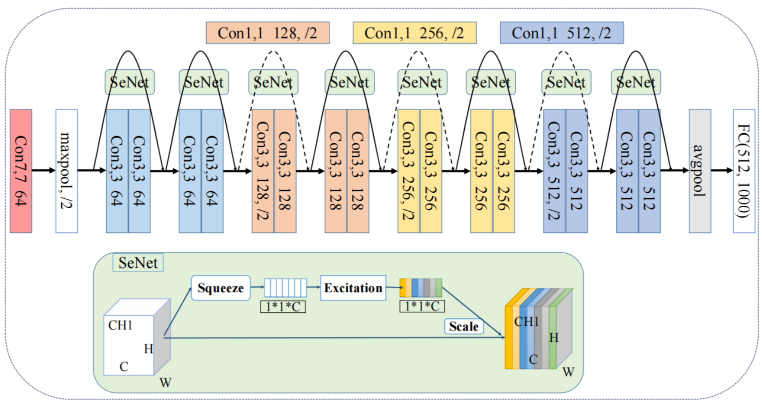

Figure 2: Model Architecture.

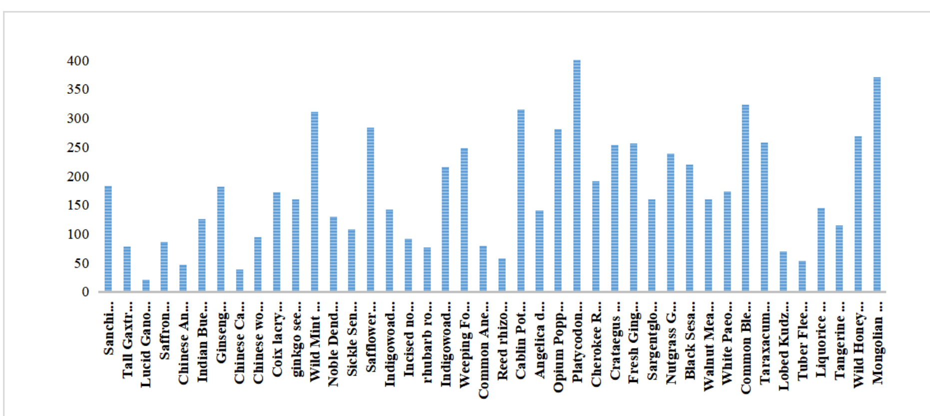

Figure 3: The Number of the Filtered Images by Using Pre-Training Models.

Assume that the dataset is D_{novel}.

$$D_{novel} = \bigcup_{n=0}^{6112} \{x_n, i\}_{i=1}^{42} \tag{3}$$

It can be seen from Figure 4 that many data in the labeling data are similar, that is, if the iterative labeling mode is not used, it is meaningless to relabel many tasks. In essence, this approach based on human-in-the-loop is to find key samples in the data for learning. Therefore, this semi-automatic labeling method has a certain degree of scientificity. Looking at the number of our labels from Figure 5, the data sample size we labeled is more than 200 for 12 groups, those with more than 100 and less than 200 are 16 groups, and those with less than 100 are 14 groups. There are 3 groups if the number is greater than 250, and 4 groups if the number is less than 50. The number distribution of the entire dataset basically meets the Gaussian distribution.

Figure 4: The Example of the CH42 Dataset.

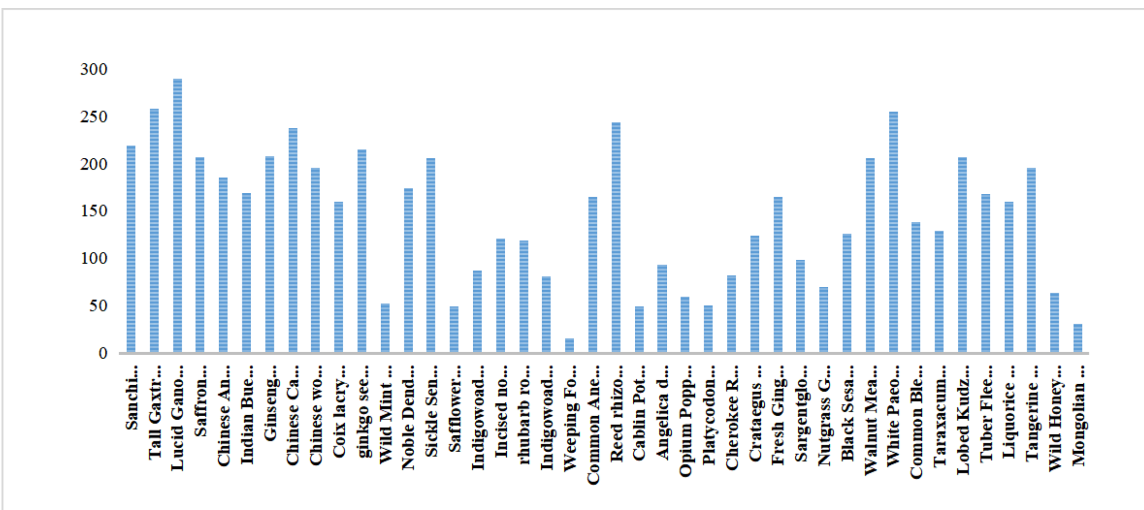

Figure 5: The Annotated Number of the CH42 Dataset.

A total of 6112 pictures were marked this time, of which 2156 were manually marked, 3,956 were automatically marked by the model, and about 64% of the data was marked by the model. This time, there are 12 sets of data with more than 200 annotations. We further checked 6 sets of pictures and found that a total of 20 pictures were marked incorrectly (the total number of pictures in the 6 sets was 1458), and the error rate was about 1.4%.

4 CONCLUSION

In this paper, we propose a method for establishing a dataset based on human-in-the-loop. As far as we know, we are the first to use the human-in-the-loop method to construct a dataset. Our framework consists of four parts: data collection, data identification, data annotation, and model update. We used this method to mark a traditional Chinese medicine dataset containing 6112 pictures, of which the model iteratively annotated 3956 pictures. Simultaneously, we sampled and checked the model annotated pictures and found that the error rate was only 1.4%. Experiments prove that our labeling method has strong robustness.

ACKNOWLEDGMENTS

This work was supported in part by the 2020 Scientific Research Fund Project of Yunnan Medical Health College (2020Y004), and the 2020 National College Students' innovation and entrepreneurship training program (202012560002).

REFERENCES

[1] Cen F, Zhao X, Li W, *et al.* (2021). Deep feature augmentation for occluded image classification[J]. *Pattern Recognition,* **111,** 107737.

[2] Wan S, Hou Y, Bao F, *et al.* (2020). Human-in-the-Loop Low-Shot Learning[J]. *IEEE Transactions on Neural Networks and Learning Systems.*

[3] Li J, Miller A H, Chopra S, *et al.* (2016). Dialogue learning with human-in-the-loop[J]. *ICLR.*

[4] Plummer B A, Kiapour M H, Zheng S, *et al.* (2019). Give me a hint! navigating image databases using human-in-the-loop feedback[C]//*2019 IEEE Winter Conference on Applications of Computer Vision (WACV).* IEEE, **2019,** 2048-2057.

[5] Böhme M, Geethal C, Pham V T (2020). Human-In-The-Loop Automatic Program Repair[C]//*2020 IEEE 13th International Conference on Software Testing, Validation and Verification (ICST)* IEEE, **2020,** 274-285.

[6] Yu F, Seff A, Zhang Y, *et al.* (2015). Lsun: Construction of a large-scale image dataset using deep learning with humans in the loop[J]. *arXiv preprint* arXiv: 1506.03365.

[7] Li G (2017). Human-in-the-loop data integration[J]. *Proceedings of the VLDB Endowment,* **10(12),** 2006-2017.

[8] Zhang S, He L, Dragut E, *et al.* (2019). How to invest my time: Lessons from human-in-the-loop entity extraction[C]//*Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining,* **2019,** 2305-2313.

[9] Klie J C, de Castilho R E, Gurevych I (2020). From zero to hero: Human-in-the-loop entity linking in low resource domains[C]//*Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics,* **2020,** 6982-6993.

[10] Yue Z, Zhang H, Sun Q, *et al.* (2020). Interventional Few-Shot Learning[J]. *Advances in Neural Information Processing Systems,* 33.

[11] Wan S, Hou Y, Bao F, *et al.* (2020). Human-in-the-Loop Low-Shot Learning[J]. *IEEE Transactions on Neural Networks and Learning Systems.*

[12] Xu Y, Wen G, Hu Y, *et al.* (2021). Multiple attentional pyramid networks for Chinese herbal recognition[J]. *Pattern Recognition,* **110,** 107558.

[13] Tóth B P, Tóth M J, Papp D, *et al.* (2016). Deep Learning and SVM Classification for Plant Recognition in Content-Based Large Scale Image Retrieval[C]//*CLEF* (Working Notes), **2016,** 569-578.

[14] Hu J, Chen Z, Yang M, *et al.* (2018). A multiscale fusion convolutional neural network for plant leaf recognition[J]. *IEEE Signal Processing Letters,* **25(6),** 853-857.

[15] Gogul I, Kumar V S (2017). Flower species recognition system using convolution neural networks and transfer learning[C]//*2017 fourth international conference on signal processing, communication and networking (ICSCN)* IEEE, **2017,** 1-6.

[16] Deng J, Dong W, Socher R, *et al.* (2009). Imagenet: A large-scale hierarchical image database[C]//*2009 IEEE conference on computer vision and pattern recognition* IEEE, **2009,** 248-255.

[17] Lin M, Chen Q, Yan S (2013 Network in network[C]. *ICLR*

[18] Simonyan K, Zisserman A (2014). Very deep convolutional networks for large-scale image recognition[C]. *ICLR.*

[19] He K, Zhang X, Ren S, *et al.* (2016). Identity mappings in deep residual networks[C]//European conference on computer vision. *Springer, Cham,* **2016,** 630-645.

[20] Zagoruyko S, Komodakis N (2016). Wide Residual Networks[C]//British Machine Vision Conference 2016. *British Machine Vision Association.*

[21] Woo S, Park J, Lee J Y, *et al.* (2018). Cbam: Convolutional block attention module[C]//*Proceedings of the European conference on computer vision (ECCV),* **2018,** 3-19.

[22] Lin T Y, Dollár P, Girshick R, *et al.* (2017). Feature pyramid networks for object detection[C]//*Proceedings of the IEEE conference on computer vision and pattern recognition,* **2017,** 2117-2125.

[23] Ioffe S, Szegedy C (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]//*International conference on machine learning. PMLR,* **2015,** 448-456.