Research for Improving the Accuracy of Image Classification Based on Semi-Supervision

Ziyang Gu[1], Lihang Wang[2] and Yueqian Zhang[3*]

[1] Institute of Science and Technology, University of Nottingham Ningbo China, Ningbo, Zhejiang, 315000, China
[2] School of Computer Engineering, Tongda College of Nanjing University of Posts and Telecommunications, Yangzhou, Jiangsu, 225127, China
[3] School of Cyber Science and Engineering, Sichuan University, Chengdu, Sichuan, 610065, China
2022141530070@stu.scu.edu.cn

Abstract. One of the core tasks of computer vision is image classification, which aims to distinguish different types of images based on various features. However, traditional image classification methods often rely on a large amount of labeled data to support them and obtaining large-scale, high-quality labeled data in practical applications is often very difficult. Insufficient data volume will affect the accuracy of image classification. In response to this issue, this article reviews semi-supervised learning methods to improve the performance of image classification. Specifically, this article selects collaborative training algorithms, self-training algorithms, and average teacher models, and applies them to the following aspects: The collaborative training strategy aims to improve the execution efficiency of image classification tasks by integrating multiple classifier algorithms and the Collaboration semi - Supervised Convolutional Neural Network (Co-S2CNN) algorithm. The self-training algorithm combines density peak and natural neighbor algorithms to reduce the weight of samples in low-density areas. The average teacher model combines the You Only Look Once (YOLO) algorithm and further introduces Strip Pooling Module (SPM) to improve the accuracy of strip object detection. Descriptive empirical research results indicate that these three hybrid algorithms can significantly improve the efficiency and accuracy of image classification.

Keywords: Image Classification, Semi-Supervision, Improving the Accuracy.

1 Introduction

The rapid advances in information technology have made image data extremely common and critical in contemporary society. Conventional image recognition techniques usually rely on a large amount of labeled data to complete the training process, which is often restricted by the cost of data acquisition and labeling in practical applications. Therefore, in current academic circles, enhancing the accuracy and efficiency of image recognition with limited label information has become a focal issue.

Y. Wang (ed.), *Proceedings of the 2024 2nd International Conference on Image, Algorithms and Artificial Intelligence (ICIAAI 2024)*, Advances in Computer Science Research 115,
https://doi.org/10.2991/978-94-6463-540-9_95

Recently, the field of image classification has benefited from the rapid progress of deep learning (DL) technology, ushering in innovative solutions. One of the key areas of DL is convolutional neural networks, which have demonstrated superior functionality in image discrimination tasks, thanks to their outstanding feature extraction and classification techniques. However, DL models usually require a large amount of labeled data for training, which is still an insurmountable problem in practical applications.

To solve this problem, this paper begins to explore the semi-supervised learning approach. This strategy lies between supervised learning and unsupervised learning and can combine a small amount of labeled data and a large amount of unlabeled data for training, to solve the challenge of scarce labeled data to a certain extent. Although semi-supervised learning has made some initial progress in image classification, there are still many research blind spots.

This study aims to explore how to effectively utilize the limited labeled samples and abundant unlabeled data to improve the classification performance in image classification tasks. This paper comprehensively discusses three semi-supervised learning methods. One is to train image data from multiple angles by fusing multiple classifiers and the Co-S2CNN algorithm, to enhance its classification efficiency. By fusing density peak technology and natural neighbor algorithm, the training mechanism reduces the importance of samples in sparse data regions. On the other hand, the average teacher model, such as combined with the YOLO family of algorithms, enhances the overall prediction accuracy. This study aims to achieve excellent classification results by using scarce labeled data, thus contributing novel ideas and technical means to the research progress of computer vision.

2 Co-training Algorithm

Co-training is a key strategy in the field of semi-supervised learning. Its basic idea is to enhance the efficiency of classifiers by using multi-view or multi-feature learning. This method uses two classifiers to simulate a small amount of labeled information. Two different classifiers infer the unlabeled data set, and select those samples with high confidence in the prediction results. The goal is to retrain the classifier by combining the unlabeled samples into the labeled data set. This process is iteratively executed until a preset termination criterion is met [1]. This strategy relies on a fundamental premise: each view or feature subset in the dataset is rich enough to independently train an efficient classifier. These classifiers cooperate to label the unlabeled information and then use the labeled information for in-depth training to improve the performance of the classifier. Traditional co-training methods are mainly limited to two classifiers. This setting weakens the efficacy of the classifier ensemble to some extent. Since only two discriminators learn from and cooperate, the information and perspective they can grasp are relatively narrow. To overcome this obstacle, researchers begin to study how to extend co-training to a wider group of classifiers, and then successfully train a new multi-classifier learning method based on co-training.

By expanding the size of the group of classifiers, multiple classifier learning algorithms can ensure that each member can absorb and integrate data from a wider range of views or features. The advantage of this method is that it is good at integrating multi-angle or multi-dimensional information in the data, which greatly enhances the accuracy and robustness of the classification. The collaboration between multiple classifiers can effectively eliminate noise and correct abnormal data, so as to significantly improve the classification efficiency. Compared with the conventional co-training algorithm, the multiple classifiers learning algorithm based on co-training shows outstanding advantages. Firstly, this method can significantly improve the classification performance in an environment of lack of labeled data. Thanks to its deep mining of valuable information in unlabeled data, and through the cooperation of multiple classifiers, it can identify and use the potential rules of data, and effectively solve the problem of lack of labeled data.

Secondly, the stability and robustness of the classification results are enhanced by integrating the efficiency of multiple classifiers. Because each classifier learns from its own unique perspective and feature level, it shows certain heterogeneity and synergy with each other [3]. As a result, when dealing with complex or uncertain data, multiple classifiers are often more accurate and stable. In addition, the multiple classifiers algorithm based on co-training shows excellent scalability. With the continuous expansion of the information base and the increase of the attributes, more classifiers can be integrated according to the actual needs, so as to effectively improve the efficiency. This adaptability enables the algorithm to meet the needs of different practical application scenarios.

In the field of image recognition, the use of a co-training algorithm has shown excellent results. As shown in Fig. 1, the Co-S2CNN algorithm integrates the powerful feature extraction ability of VGGNet, GoogLeNet, ResNet, and other models, as well as the potential information of unlabeled data, and achieves a significant improvement in the efficiency of image classification in the context of limited labeled samples [5]. This method uses co-training and pseudo-label generation technology to jointly train many convolutional neural network models. Through the idea of disagreement generation and co-training, the low-efficiency classifier is improved into a more efficient strong classifier. As can be seen from Table 1, compared with the single Convolutional Neural Network (CNN) model, the Co-S2CNN algorithm shows significant superiority, and its accuracy is also significantly improved.

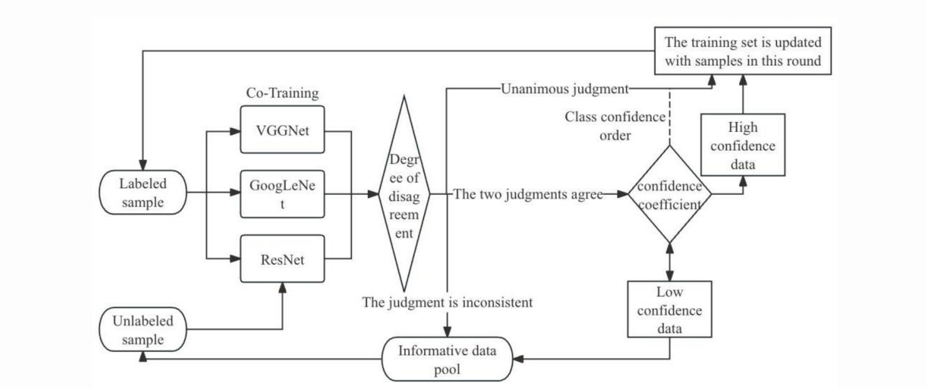

Fig. 1. Co-training process of Co-S2CNN algorithm [5]

Table 1. Model accuracy comparison [5]

	The initial tag set is a single-model			Co－S2CNN algorithm
Caltech－UCSD Birds－200－2011	0.6277	0.6475	0.6382	0.7970
Caltech－256	0.7883	0.7561	0.7386	0.8993

Compared with the traditional single classifier or simple dual classifier for co-training, they often show certain shortcomings in dealing with complex image recognition tasks. They may not be able to dig out the multi-view or multi-feature information in the data, and they may not be able to deal with noise and outliers in the data efficiently. The multi-classifier algorithm based on co-training effectively solves these problems. It enhances the accuracy and stability of classification by fusing multiple classifiers for co-training [4]. It can be seen that in the field of semi-supervised learning, the multiple-classifier algorithm relying on collaborative training shows outstanding advantages. By broadening the co-training algorithm to a larger ensemble of classifiers, the efficiency of co-training can be greatly improved, which makes it possible to achieve excellent classification results in environments where labeled data is scarce.

3 Self-training Algorithm

Self training algorithms, as a classic SSL algorithm, have achieved significant development and results. The algorithm was pre-trained on unlabeled datasets and then fine-tuned with a small amount of labeled data to improve the model's performance. However, with the diversity and complexity of the data, there are still two main issues. (1) When new class samples appear in the training dataset, the performance of the algorithm will be affected because the labels of the new class samples are unknown during the training process. (2) The manually annotated method of obtaining labeled samples requires the participation of domain experts, consumes high time and financial costs, and carries the risk of human error labeling. Researchers can use some modifications to improve algorithms, such as combining density peak and natural neighbor algorithms. The density peak algorithm refers to using the density information of the density peak clustering algorithm to calculate the outlier index of samples. Specifically, in this algorithm, a Gaussian kernel function is used to define the density of the sample, and then obtain its local density by calculating the distance between the sample and other samples [6].

The natural neighbor algorithm is a scale-free nearest neighbor algorithm that can automatically form neighbor relationships without manually setting parameters. For each sample xi, the natural neighbor algorithm will find the sample xj that is closest to each other, treating xj as xi's natural neighbor. The search process for natural neighbors starts with k=1, gradually searching for the first k nearest neighbors, and recording the number of times each sample appears among the k nearest neighbors of other samples. Then, increase the value of k and repeat the above steps until each sample has natural neighbors, or there are no more samples without natural neighbors. The core idea of the natural neighbor algorithm is to construct neighbor relationships by finding samples that are closest to the target sample, without the need to pre-set the number or distance threshold of neighbors. Set sample weights based on global density and outlier index, and remove outliers. This improved method can reduce the weight of samples in low-density areas and improve classification accuracy. By combining density peaks and natural neighbors, this algorithm can solve the problem of self-training algorithms discovering high-confidence unlabeled samples and incorrectly labeled samples during the training process. The following table lists five self-training combined methods which are the Peak density self-training algorithm(STDP), combining peak density and the nearest neighbor noise filter self-training algorithm(STDPENN), combining peak density and all nearest neighbor noise filter self-training algorithm(STDPAKNN), combining peak density and edge cutting noise filter self-training algorithm(STDPCEWS), combining peak density and natural neighbor noise filter self-training algorithm(STDPNNN).and also the table shows their classification accuracy in different fields.

Table 2. Compares the accuracy and standard deviation of various algorithms.

Data set	STDP	STDPENN	STDPAKNN	STDPCEWS	STDPNNN
Audio	86.46%±2.94%	86.20%±3.11%	85.95%±3.01%	86.21%±3.10%	86.46%±2.94%
Dermatology	71.83%±8.75%	72.08%±8.81%	72.37%±8.62%	72.06%±8.55%	71.82%±8.60%
Glass	81.10%±8.45%	81.16%±8.75%	81.23%±8.68%	81.20%±8.75%	81.38%±7.94%
Image Segmentation	87.84%±2.44%	87.92%±2.60%	88.06%±2.65%	87.96%±2.66%	87.36%±2.51%
Indian Liver Patient Dataset	71.19%±7.15%	71.01%±4.96%	70.16%±4.08%	71.06%±4.97%	71.01%±7.23%
Average value	76.82%	77.33%	76.79%	77.35%	77.53%

By using a total of 14 statistical tables in Table 2 (some of which are listed here), the average value can be obtained [7]. It can be seen that the combined density peak and natural neighbor algorithm can achieve higher accuracy than other algorithms in the case of fewer sample labels. These UCI data include various fields such as audio, dermatology, glass, image segmentation, and liver diseases. So this self-training algorithm can also be well applied to solve the segmentation of cardiac medical images. There are problems in this field, such as the relatively small amount of data to be processed and the requirement for high accuracy of segmentation results; for annotation of medical images, professional medical knowledge is required, and this process itself involves complexity and time-consuming characteristics. Due to the diversity and complexity of medical images, performing well on the training set does not necessarily mean achieving the same performance on the testing set [8]. However, the self-training combined algorithm introduced in this article can be applied to this problem, not only by expanding the dataset through self-training to solve the problem of data scarcity but also by continuously iterating and optimizing the algorithm. This model can gradually grasp the basic laws of the target data and improve its generalization ability and segmentation accuracy.

4 Mean Teacher

4.1 Semi-Supervised Algorithm

Mean TeacherIn semi-supervised learning, unlabeled data does not have labels and categories, and consistency regularization methods do not require label information. That is, if there is a slight disturbance in the input that does not affect the prediction, it is consistent [9]. In the specific use of semi-supervised learning, consistency regularization is used on unlabeled samples, while cross-entropy is calculated for errors on labeled samples as originally done. Finally, the two parts of the loss are combined to calculate the overall loss function. Among them, the teacher-student model based on consistency regularization is based on the original π model and the temporal ensemble model, mainly developing on the temporal ensemble model. In the temporal ensemble model, only the unsupervised consistency input is observed. The first time is the image parameters of data augmentation. The second method uses the exponential moving average (EMA) of iterative prediction results but only once per iteration. The teacher-student model, as shown in Fig. 2, employs EMA multiple times. In the Mean Teacher model, noise is added to both the student and teacher layers, and consistency constraints are applied to both models. Labeled samples are input into the student model to calculate cross-entropy loss 1. Unlabeled samples are then input into both the student and teacher models for training and the output of the student model is obtained. Then, the mean squared error of the two results obtained after training the teacher and student models with unlabeled samples is calculated to obtain the loss 2. Finally, the total loss is calculated by combining the losses 1 and 2 [10]. In this process, the student model's weights update the teacher model's weights through EMA with each update. This realizes the use of EMA to incorporate new parameters into the model for updates in each iteration. The student model also utilizes the labels trained by the teacher, so that the student and teacher models train each other, which can better integrate new information and achieve more efficient training.

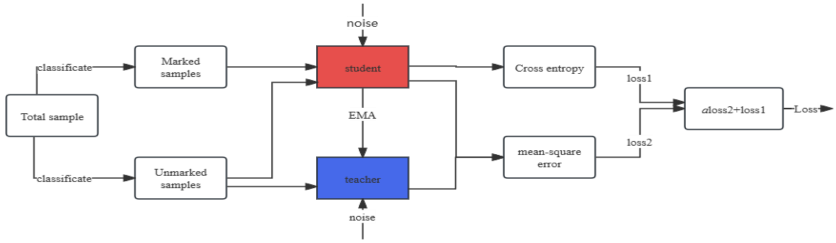

Fig. 2. Average Teacher Model Process

4.2 Object Detection Algorithm

YOLO Algorithm Object detection algorithms are often used for classifying images, but they rely heavily on the number of labeled samples. In reality, labeling samples consumes a significant amount of labor and resources, so semi-supervised algorithms can be introduced to reduce this dependency.

Object detection algorithms can be divided into two categories: one-stage and two-stage. One-stage algorithms directly perform the regression task, which is faster than two-stage algorithms that first segment the images into multiple bounding boxes before outputting through a convolutional neural network. The YOLO series of algorithms is a model based on deep neural networks for object detection within the one-stage category, known for its fast speed and high accuracy. However, this model is still very dependent on labeled samples. Therefore, a semi-supervised teacher-student model method is introduced here.

4.3 The Method Combining the Two Approaches

In practice, an SPM - YOLOv5 algorithm based on the teacher-student model has made significant progress in the identification of bagged citrus fruits. The algorithm first acquired many images of citrus fruits from orchards, then performed data augmentation, and after partially labeling the data, used the algorithm for detection. By incorporating the SPM module into the original YOLOv5 algorithm, the accuracy of strip-shaped object detection was improved, making it suitable for the strip-shaped changes in the shape of citrus fruits after bagging. Then, by integrating the semi-supervised teacher-student model, the dependency on labeled samples was reduced. The teacher model was initially trained on a dataset with annotated samples. In the training of the student model, labeled samples were trained directly, while unlabeled samples relied on the teacher model for updates based on the training results; in the training of the teacher model, the samples trained by the student model were used to update and adjust the teacher model in each iteration through EMA (Exponential Moving Average), ultimately resulting in a model with relatively high accuracy [11].

This method has achieved a significant improvement in average accuracy in the identification tasks of citrus and branches where bagging technology is applied. As shown in Table 3, to highlight the improvement in accuracy by combining semi-supervised methods, only the teacher-student model + YOLOv5 is compared with YOLOv5 here, excluding the impact of SPM (which emphasizes the detection of strip-shaped objects, such as bagged citrus in this case). A 5.8% increase in accuracy was achieved for branches and bagged citrus with branches, as well as for bagged citrus alone, and an increase of 7.2% in average accuracy for branches [11].

However, this method is only used in agriculture, and it adds the SPM to the YOLO algorithm, making it more suitable for detecting strip-shaped objects and limiting its practicality in detecting other objects [12]. However, this method can be applied beyond agriculture. By using a higher version of the YOLO algorithm based on the teacher-student model, such as YOLOv8, and removing the SPM, it can be more widely applied to various image and graphic recognition tasks, thereby enhancing recognition accuracy.

Table 3. Comparison of experimental results between two models [10]

Models for object detection	Backbone	Model size/MB	Class	Mean Average Precision mAP/%	Precision/%	Detection speed/(frame·s-1)
YOLOv5	YOLOv5s	20.3	Bagged citrus+branches	57.9	79.4	26
			Bagged citrus	69.9	89.4	26
			branches	46.0	69.3	26
Teacher student model+YOLOv5	YOLOv5s	55.5	Bagged citrus+branches	59.7	85.2	26
			Bagged citrus	66.1	95.2	26
			branches	53.2	75.1	26

In other areas, the semi-supervised Mean Teacher model also has many enhancing effects on image classification:

Without modifying its underlying network structure, the model known as II by Laine & Aila within TensorFlow served as the baseline. This model is based on a 13-layer convolutional neural network architecture. Experiments were performed on two comprehensive datasets: the Street View House Numbers (SVHN) and the CIFAR-10, both of which encompass a diverse array of images. The SVHN dataset focuses on close-up images of house numbers, while the CIFAR-10 includes a variety of natural images such as cats, dogs, airplanes, and cars. Upon integrating the semi-supervised Mean Teacher model, the SVHN dataset, with a limited set of only 250 labeled images, achieved a 4.35% error rate. Furthermore, when the Mean Teacher model was paired with a residual network on the CIFAR-10 dataset, the error rate was significantly lowered from 10.55% to 6.28%. Similarly, on the ImageNet 2012 dataset with just 10% of the labels available, the error rate was dramatically reduced from 35.24% to 9.11% [13].

The classification of rock thin-section images often relies on labeled data, which is more challenging to obtain and also depends on the expertise and technical skills of the classifiers. Therefore, based on the pre-trained VGG16 model, not only was the Mean

Teacher model used, but it was also improved by incorporating a layer consistency regularization method, resulting in the Layer Consistency Mean Teacher model. After comparative experiments, this model demonstrated that when using semi-labeled datasets of varying degrees, compared with the pre-trained VGG16 model using fully labeled datasets with different sampling ratios (all datasets coming from rock thin-section images on the Scientific Data Bank website), the classification ability of the model with a 50% semi-labeled dataset was almost as good as that of the pre-trained VGG16 model with a 100% labeled dataset. Moreover, when using datasets with labeling ratios of 30%, 50%, and 70%, the Layer Consistency Mean Teacher model achieved accuracy improvements of 10.7%, 8.5%, and 6.4% over the VGG16 model, respectively [2].

5 Conclusion

Currently, supervised algorithms have yielded numerous outcomes across the spectrum of image recognition applications. However, these algorithms are constrained by the limitations of available labeled data. Hybrid algorithms that integrate semi-supervised learning with other methodologies have demonstrated pronounced benefits within the domain of image recognition. This paper, therefore, discusses three distinct semi-supervised learning algorithms in conjunction with other algorithms, contrasting their performance with that of semi-supervised algorithms used solely for recognition tasks. Co-training strategies bolster the classifier's performance through multi-perspective learning, while self-training algorithms, in tandem with density peak and natural neighbor methods, effectively mitigate the effects of noise. Additionally, the Mean Teacher model enhances the model's generalization capabilities via consistency regularization. Comparative analysis indicates that these hybrid algorithms, by leveraging a modest amount of labeled data alongside a vast pool of unlabeled data, have not only elevated the model's training efficiency but also markedly improved the accuracy of recognition. The successful implementation of these hybrid algorithms introduces new opportunities for achieving higher precision in image recognition endeavors, significantly contributing to the advancement of computer vision technologies. Despite the promising outcomes of the algorithms presented in this paper within the scope of image classification tasks, there exists scope for further refinement. Subsequent research endeavors could delve into an expanded array of data augmentation techniques, refine algorithmic structures, and aim to maximize efficiency with minimal reliance on labeled tags. Validation across an even broader spectrum of datasets could also enhance the robustness of these findings. Researchers are encouraged to extend their exploration to the application of these algorithms in diverse domains, including but not limited to video analysis and natural language processing. It is crucial to acknowledge that while hybrid algorithms have the potential to elevate classification performance, they may concurrently introduce increased complexity and computational demands on the model. Future endeavors must strive to balance the imperative for precision with considerations of algorithmic scalability and real-time applicability.

Authors Contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

1. Wang, W., Zhou, Z. H.: Analyzing co-training style algorithms. Springer-Verlag (2007)
2. Yan, Z. J.: Semi-supervised rock slice image classification based on layer consistent average teacher model. Journal of Applied Science 42(1): 27-38 (2024)
3. Solyman, M., Gayar, N. E.: A co-training approach for semi-supervised multiple classifier (2024)
4. Lu, J. L.: Comparison and improvement of collaborative training and multi-view methods in semi-supervised learning. Ocean University of China (2012)
5. Ge, M. Y., Yu, C. C., Zou, L., et al.: DL Image classification algorithm based on collaborative semi-supervised. Computer Simulation (2), 5 (2019)
6. He, Y. L., Chen, J. Q., Huang, Q. H. et al.: Semi-supervised learning algorithm for self-training novel class detection. Shenzhen University. School of Computer and Software, Guangdong Laboratory of Artificial Intelligence and Digital Economy (Shenzhen). 6 (2022)
7. Lv, J., Liu, Q., Li, S. J.: Self-training algorithm combining density peaks and improved natural neighbors. Journal of Nanjing University (Natural Sciences), 9 (2022)
8. Li, Y. B.: Research on cardiac medical image segmentation algorithm based on self-attention mechanism and self-training method. Shandong University. School of Mechanical and Information Engineering. 8 (2023)
9. Ciresan, D., Meier, U., Gambardella, L., et al.: Deep, big, simple neural nets for handwritten digit recognition. Neural Computation 22(12): 3207-3220 (2010)
10. Lv, H. Y., Yu, L., Zhou, X. Y., et al.: Survey of semi-supervised DL Image Classification Methods. Journal of Frontiers of Computer Science & Technology 15(6) (2023)
11. Lv, J., Li, S. J., Zeng, M. Y., et al.: Bagged citrus detection algorithm based on semi-supervised SPM-YOLOv5. Transactions of the Chinese Society of Agricultural Engineering 38(18): 204-211 (2022)
12. Hou, Q., Zhang, L., Cheng, M. M., et al.: Strip pooling: Rethinking spatial pooling for scene parsing. Seattle, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 4003-4012 (2020)
13. Tarvainen, A., Harri, V.: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised DL results. Advances in neural information processing systems 30 (2017)