

Article

ANN-Based Continual Classification in Agriculture

Yang Li ^{1,2,*}  and Xuewei Chao ¹ 

¹ College of Mechanical and Electrical Engineering, Shihezi University, Xinjiang 832003, China; psicnhpu@163.com

² School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China

* Correspondence: liyang328@shzu.edu.cn

Received: 13 April 2020; Accepted: 15 May 2020; Published: 18 May 2020



Abstract: In the area of plant protection and precision farming, timely detection and classification of plant diseases and crop pests play crucial roles in the management and decision-making. Recently, there have been many artificial neural network (ANN) methods used in agricultural classification tasks, which are task specific and require big datasets. These two characteristics are quite different from how humans learn intelligently. Undoubtedly, it would be exciting if the models can accumulate knowledge to handle continual tasks. Towards this goal, we propose an ANN-based continual classification method via memory storage and retrieval, with two clear advantages: Few data and high flexibility. This proposed ANN-based model combines a convolutional neural network (CNN) and generative adversarial network (GAN). Through learning of the similarity between input paired data, the CNN part only requires few raw data to achieve a good performance, suitable for a classification task. The GAN part is used to extract important information from old tasks and generate abstracted images as memory for the future task. Experimental results show that the regular CNN model performs poorly on the continual tasks (pest and plant classification), due to the forgetting problem. However, our proposed method can distinguish all the categories from new and old tasks with good performance, owing to its ability of accumulating knowledge and alleviating forgetting. There are so many possible applications of this proposed approach in the agricultural field, for instance, the intelligent fruit picking robots, which can recognize and pick different kinds of fruits; the plant protection is achieved by automatic identification of diseases and pests, which can continuously improve the detection range. Thus, this work also provides a reference for other studies towards more intelligent and flexible applications in agriculture.

Keywords: similarity; metric; memory; deep learning

1. Introduction

In the field of intelligent agriculture, for instance, plant protection and precision farming, there are incremental progresses in agricultural image processing, e.g., classification of crop pests, and harvest yield forecast. Step advances are catalyzed by the developed various computerized models, which have covered a wide range of technologies, such as machine learning, deep learning, transfer learning, few-shot learning, and so on. For instance, several machine learning methods were adopted in crop pest classification [1,2]. The convolutional neural networks were used to diagnose and identify the plant diseases from leaf images [3,4]. The deep learning neural networks showed a powerful and excellent performance on several agricultural applications, such as plant identification [5], crop classification [6,7], fruit classification [8], weed classification [9], animal classification [10], quality evaluation [11], and field pest classification [12,13]. The transfer learning technology helped fine-tune the pre-trained models to reduce the difficulty of model training [14,15]. The few-shot learning method reduced the requirements for the scale of the training dataset [16]. There were also some related agricultural research surveys [17–19], providing more comprehensive views.

Although the abovementioned methods achieved good performance on some special tasks, they are still far away from the true intelligence in this area. Specifically, one deep neural network is designed for a special task with a static evaluation protocol. The entire dataset will be split in two parts: A training set used for learning and a testing set used for accuracy evaluation. Once the training period completes, the structure and parameters of this model are fixed, and any new knowledge cannot be learned again. This is quite different from how humans learn.

Biological learning is to continually learn new skills (tasks) and accumulate knowledge throughout the lifetime [20]. We can also incorporate new information to expand our cognitive abilities without seriously breaking past memories, which results from the good balance between synaptic plasticity and stability [21,22]. As known, the basic principle of deep learning networks is the back-propagation error and gradient descent. However, from the perspective of biological cognition, our learning process is more likely through similarity matching, rather than back-propagation error or gradient descent in the brain [23]. So, the bio-inspired work in this article will be around metric learning and continual learning. The metric learning aims to learn the internal similarity between input paired data [24], which is suitable for classification and pattern recognition in agriculture. Continual learning requires the designed model to continuously learn new tasks without forgetting old ones, that is, keeping good performance on both new and old tasks.

Continual learning is an approach inspired from the biological factors of the mammalian brain. In this topic, the most important thing is the stability–plasticity dilemma. In detail, plasticity means to integrate novel information to incrementally refine and transfer knowledge, and stability aims not to catastrophically interfere with consolidated knowledge. For a stable continual learning process, two types of plasticity are required: Hebbian plasticity for positive feedback instability and compensatory homeostatic plasticity, which stabilizes neural activity [25]. So far, the main methods to realize continual learning can be divided into three categories: (1) Store and replay, including previous data or memory. The limitation is that the storage of old information will lead to large working memory requirements; (2) regularization approaches, such as learning without forgetting (LWF), elastic weight consolidation (EWC), etc. They alleviate catastrophic forgetting by imposing constraints on the update of the neural weights. However, the additional loss terms for protecting consolidated knowledge will lead to a trade-off on the performance of old and novel tasks; and (3) dynamic architectures, which change the structure of networks in response to new information, e.g., re-training with an increased number of neurons or network layers. Obviously, the limitation is the continuously growing complexity of the architecture with the increasing number of learnt tasks.

In this study, in order to imitate human learning and memory patterns to maintain a good performance on both new and old tasks, we propose an artificial neural network (ANN)-based continual classification method via memory storage and retrieval, including the convolutional neural network (CNN) and generative adversarial network (GAN). Looking at ourselves, how do we remember past events? We only keep the most important information in our brain, throwing out the details and abstracting the inner relationships. These life experiences inspire us to find a way to abstract and preserve prior knowledge in memory. The memory only records the most important information from prior events, automatically ignoring the details. Inspired by this, we used the GAN to extract central information from old tasks and generate abstracted images as memory. For the similarity matching tasks in agriculture, it has a good effect on both new and old tasks, alleviating the forgetting problem. The main contributions of this work are as follows:

- (1) The CNN model only requires few raw data, which is helpful for practical applications, such as classification and pattern recognition.
- (2) The proposed method based on memory storage and retrieval can deal with the sequential tasks and maintain a good performance on both new and old tasks, without forgetting.
- (3) This work has two important advantages: Few data and high flexibility. It provides a foundation for other relevant studies toward more flexible applications in agriculture.

Clearly, there are so many possible applications of this proposed approach in the field of agriculture, for instance, intelligent fruit picking robots, which can recognize and pick different kinds of fruits, and plant protection through automatic identification of diseases and pests, which can continuously improve the detection range to show the ability to upgrade the developed model.

2. Materials and Methods

2.1. Crop Pest and Plant Leaf Datasets

The typical deep neural networks require amounts of data to train the model, while the metric learning method only needs few raw data. In this research work, we collected two common cross-domain agricultural datasets: Crop pests and plant leaves. The crop pest dataset was collected from the open dataset [26], which provides images of important insects in agriculture with natural scenes and complex backgrounds, close to the real world. The plant leaf dataset was collected from the famous open dataset (PlantVillage). Generally, the image preprocessing for deep neural networks includes the target cropping, background operation, gray transform, etc. Here, we used the raw images to make this study closer to the practical application.

In the crop pest dataset, there are 10 categories and the number of samples in each class is 20. The total number of samples is 200, which is a small dataset compared to that required for the traditional deep learning models based on back-propagation error. Some samples of the crop pest dataset are shown in Figure 1.

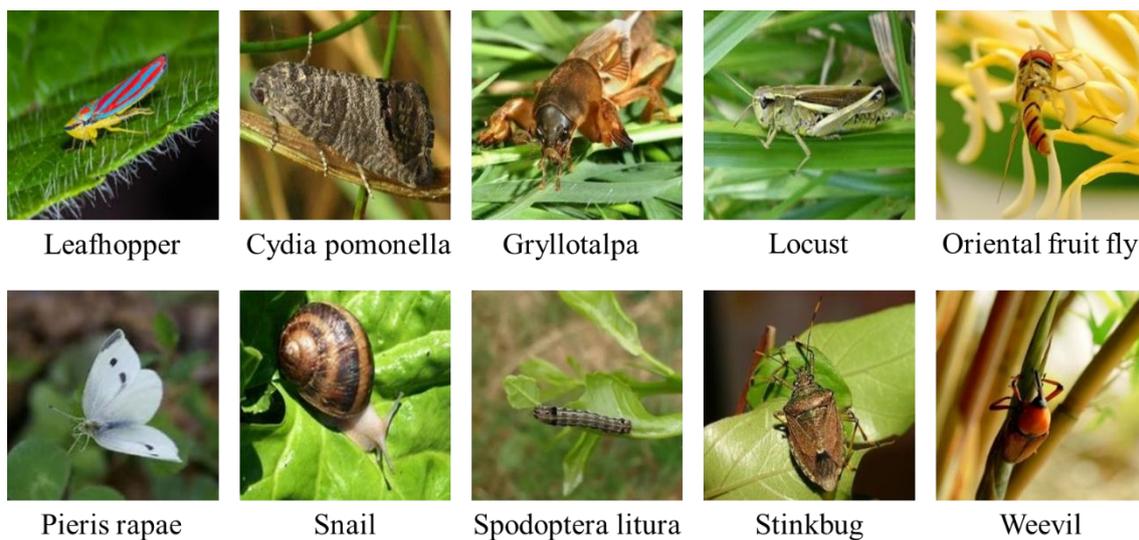


Figure 1. Samples of the crop pest dataset (from [26]).

The plant leaf dataset also includes 10 classes, and the number of samples in each class is 20. The parameter sizes of these two databases are the same. Some samples of the plant leaf dataset are shown in Figure 2.

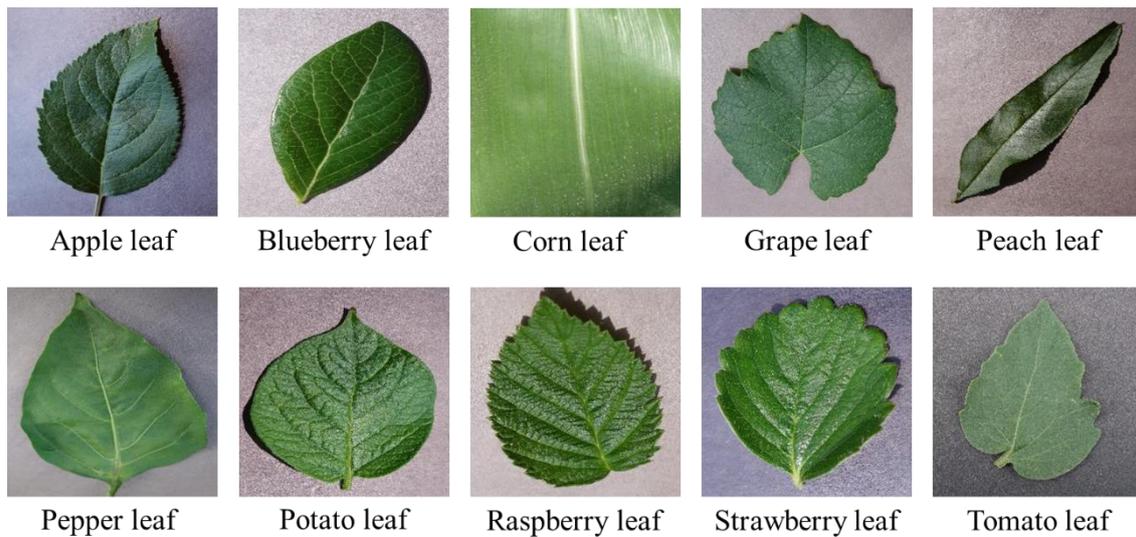


Figure 2. Samples of the plant leaf dataset (from PlantVillage).

2.2. Classification with Metric Learning Based on CNN

Metric learning learns the inner similarity between input paired data using a distance metric, which is aimed at distinguishing and classifying. The typical metric learning model is the Siamese network [27]. The Siamese network basically consists of two symmetrical neural networks sharing the same weights and architecture, which are joined together at the end using some energy function. During the training period of the Siamese network, the inputs are a pair of images, and the objective is to distinguish whether the input paired images are similar or dissimilar. The workflow of the Siamese network is shown as Figure 3.

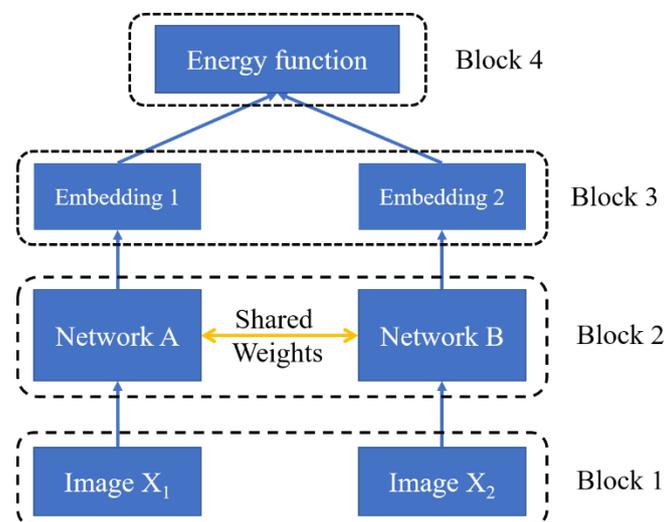


Figure 3. The workflow of the Siamese network.

As shown in Figure 3, there are four blocks. Now, we considered them one by one. For block 1, it means the input paired images, including the images X_1 and X_2 , fed to network A and network B, respectively. They may come from the same category or not.

For block 2, there are two convolutional neural networks (CNNs), named network A and network B. The role of network A and network B is to generate the embeddings (feature vectors) for the input paired images. Since the inputs of the model are images, we used a CNN to generate the embeddings. Remember that the role of the CNNs here is only to extract features but not to classify. This differs with

the traditional deep learning classification models. It is required that the two CNNs in the Siamese network have shared weights and structure, which means the two CNNs, in fact, have the same topology, as shown in Figure 4.

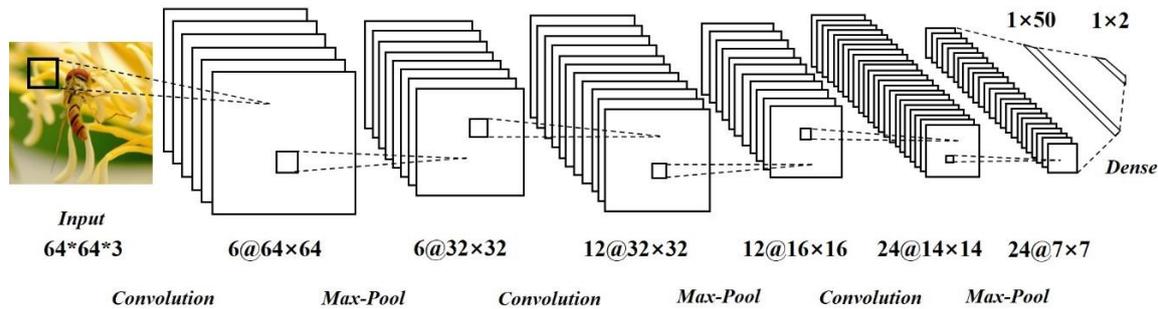


Figure 4. The topology of used convolutional neural network (CNN).

Here, the shared structure and parameters of CNN are shown in Table 1. Specifically, the output shape of the layers in CNN, and the size and number of kernels used in the convolutional layers and max-pooling layers are included. The programming tool used was ‘Jupyter Notebook’, which is a popular web-based interactive computing environment. We realized the functions with Python language and the environmental backend was TensorFlow. Our programming files and used image dataset were uploaded to the ZENODO.org, which is free and open for other researchers [28].

Table 1. The CNN structure and parameters in the Siamese network.

Layers	Output Shape	Kernel Size	Kernel Num	Stride	Parameters
Input	(64, 64, 3)				0
Conv2D	(64, 64, 6)	(3, 3)	6	1	168
MaxPooling2D	(32, 32, 6)	(2, 2)	0	2	0
Conv2D	(32, 32, 12)	(3, 3)	12	1	660
MaxPooling2D	(16, 16, 12)	(2, 2)	0	2	0
Conv2D	(14, 14, 24)	(3, 3)	24	1	2616
MaxPooling2D	(7, 7, 24)	(2, 2)	0	2	0
Dense	(50)				58,850
Dense	(2)				102

Then, for block 3, the embedding is referred to the output of the last dense layer of CNN, as shown in Figure 4. Network A and network B generate the embeddings for the input images X_1 and X_2 , respectively. These embeddings are fed to block 4, the energy function, which gives the similarity between the paired inputs. The Euclidean distance is adopted as the energy function, which is the most common way to measure the distance between the two embeddings in the high-dimensional space. The expression of block 4, the energy function, can be written as Equation (1):

$$E(X_1, X_2) = \|f_{N_A}(X_1) - f_{N_B}(X_2)\|_2 \quad (1)$$

The value of E represents the similarity between the outputs of the two networks: If X_1 and X_2 are similar (from the same category), the value of E will be less. Otherwise, the value of E will be large if the inputs are dissimilar (from different categories).

To train the Siamese network well, the loss function is very important. The loss function guides the iteration of parameters of CNNs in the Siamese network. Since the goal of the Siamese network is to understand the similarity between the paired input images, we used the contrastive loss function, expressed as Equation (2):

$$\text{Contrastive Loss} = Y \times E^2 + (1 - Y) \times [\max(\text{margin} - E, 0)]^2 \quad (2)$$

where E is the energy function and Y is the true label, which is 0 if the two input images are from the same category and 1 if the two input images are from different categories. Some examples of the input pairs are shown in Figure 5.

Input pairs		Label
		0
		1

Figure 5. Examples of the input pairs.

In Equation (2), the term margin is used to set the threshold, that is, when input pairs are dissimilar, the Siamese network needs to hold their distance greater than the margin; otherwise, there will be a loss during the training period. Here, the margin was set as 1. When the training period is done, the distribution of embeddings will have a group effect, where different groups represent different categories.

2.3. Continual Classification with Metric Learning Based on CNN and GAN

From the bio-inspired perspective, we aimed for the model to be more flexible and able to handle continual tasks. Continual learning, also called lifelong learning, differs from transfer learning or other traditional networks. As known, a typical deep neural network is designed for some specific task, e.g., crop pest classification. After the training period, the weights and structure of the designed model are fixed, with an excellent performance on the specific task. However, if we want the model to perform another new task directly, e.g., plant leaf classification, it will have a very bad performance unless it is trained again from scratch or uses transfer learning. However, if we train the model by the new dataset, the distribution of weights will change to ensure a good performance on a new task. Since the weights of the network are modified, the network loses the ability to recognize the old task; in other words, it forgets the old knowledge. For transfer learning, the forgetting problem of old knowledges still exists. Obviously, traditional learning way has very poor flexibility.

If we want a model that can continually learn new tasks without forgetting old knowledge, it should have some bio-inspired ability, such as memory. In this study, we proposed a continual classification method based on memory storage and retrieval to maintain a good performance on both new and old tasks. Look at ourselves, how do we remember past events? We only keep the most important information in our brain, throwing out the details and abstracting the inner relationships. These life experiences inspire us to find a way to abstract and preserve prior knowledge in memory.

Here, we used the GAN to perform information abstracting and memory storage, which is a technique to learn to generate new data with the same statistics as the raw dataset, which consisted of two parts: Generator and discriminator. The basic workflow of GAN is shown as Figure 6.

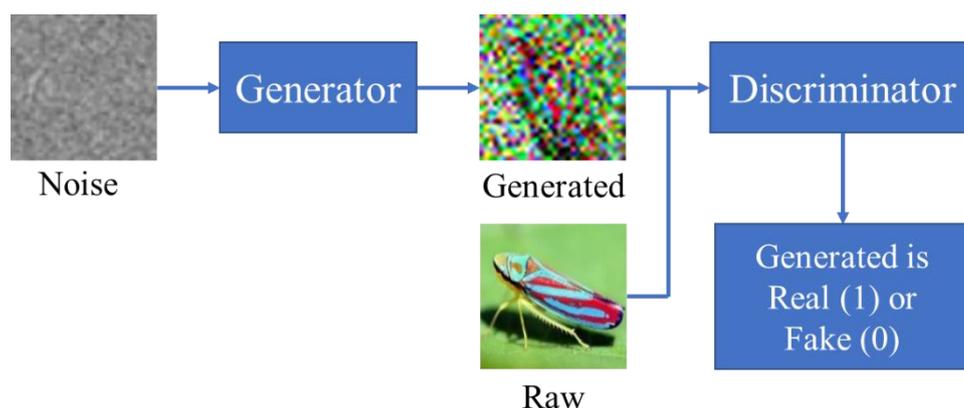


Figure 6. The workflow of generative adversarial network (GAN).

The generator and discriminator are both deep convolutional neural networks, and their structures are shown in Table 2.

Table 2. The generator and discriminator in GAN.

Generator		Discriminator	
Layers	Output Shape	Layers	Output Shape
Input	(64)	Input	(64, 64, 3)
Dense	(32, 32, 128)	Conv2D	(62, 62, 128)
Conv2D	(32, 32, 256)	Conv2D	(30, 30, 128)
Conv2DTranspose	(64, 64, 256)	Conv2D	(14, 14, 128)
Conv2D	(64, 64, 256)	Conv2D	(6, 6, 128)
Conv2D	(64, 64, 3)	Dense	(1)

The GAN chains the generator and discriminator together, expressed as Equation (3):

$$GAN(X) = discriminator(generator(X)) \quad (3)$$

The generator and discriminator contest with each other in a game. We trained the discriminator using samples of raw and generated images with the corresponding labels, such as any regular image classification model. To train the generator, we started with the random noise and used the gradients of the generator's weights, which means, at every step, moving the weights of the generator in a direction that will make the discriminator more likely to classify the images decoded by the generator as "real". In other words, we trained the generator to fool the discriminator.

Since the GAN can carry out the memory storage for old tasks, the workflow of our proposed continual metric learning method can be shown as Figure 7, which is mainly based on memory storage and retrieval.

When the first task comes, the task data will be organized as pairs and fed to the metric learning model (Siamese network). The output result is the similarity between input pairs, that is to say whether the input images are from the same category or not. Besides, the task data will also be fed to the GAN after data augmentation, due to the small scale of the raw database. Then, the GAN generates the abstracted images that represent the most important information of the old tasks, after the amount of iterations. We call this process memory storage. When the second task comes, the new task data and the data from memory will be mixed together, and fed to the metric learning model. We call this process memory retrieval.

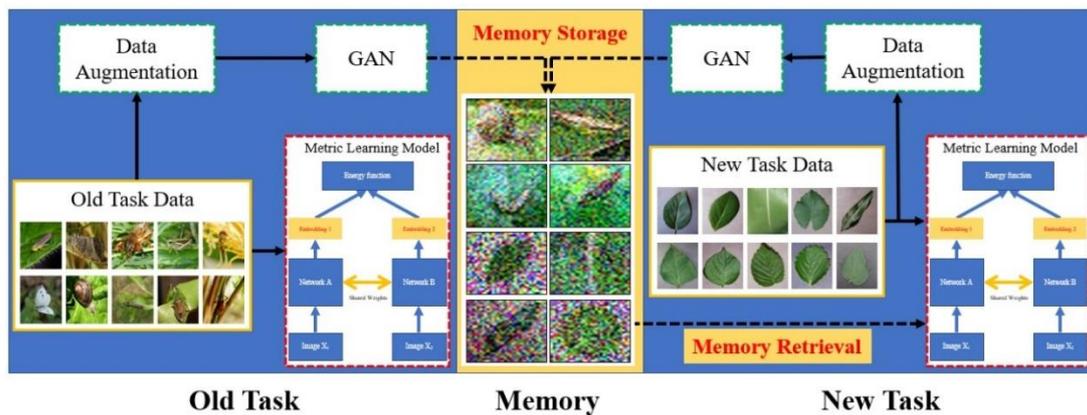


Figure 7. The workflow of continual metric learning.

3. Results

3.1. Single Task Experiment with the Basic CNN Model

In order to testify the performance of the metric learning model on similarity matching for a single task, we carried out experiments on a crop pest dataset and plant leaf dataset, respectively. For these two datasets, we prepared the input data as paired images. In detail, the total number of input pairs was 10,000, which may have contained a small number of duplicates because of the random combinations. We split the training set and testing set by the ratio of 8:2, that is, 2000 input pairs were used to test the accuracy. During training, 25% of the training data were taken out for the validation set. In summary, there were 6000 pairs for training, 2000 pairs for validation, and 2000 pairs for testing.

For the crop pest dataset, the loss and accuracy of the CNN model is shown in Figure 8.

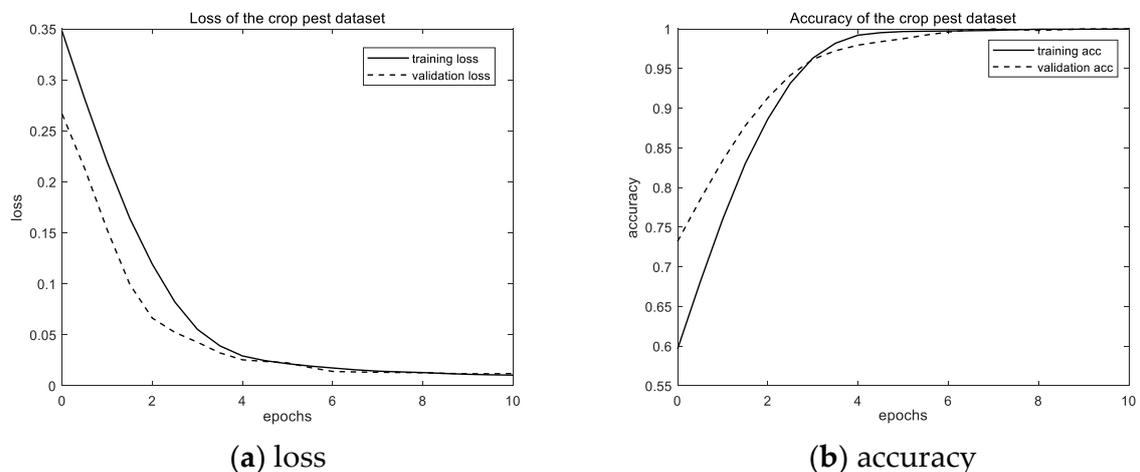


Figure 8. The loss and accuracy on the crop pest dataset.

It is shown that the variation trend of the training loss is consistent with that of the validation loss. The variation trend of the training accuracy is also consistent with that of the validation accuracy. This indicates that there is no overfitting problem in the training. The testing accuracy is 100%, which means the model can distinguish the input paired images well. The distribution of embeddings from the crop pest dataset is shown in Figure 9.

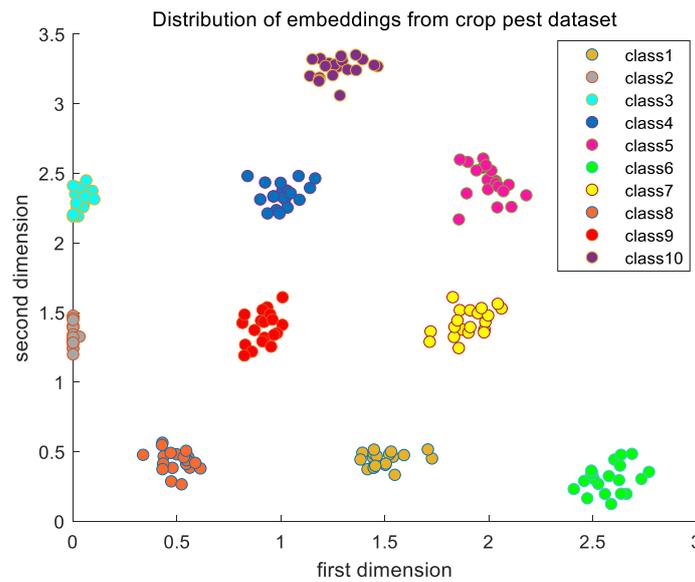
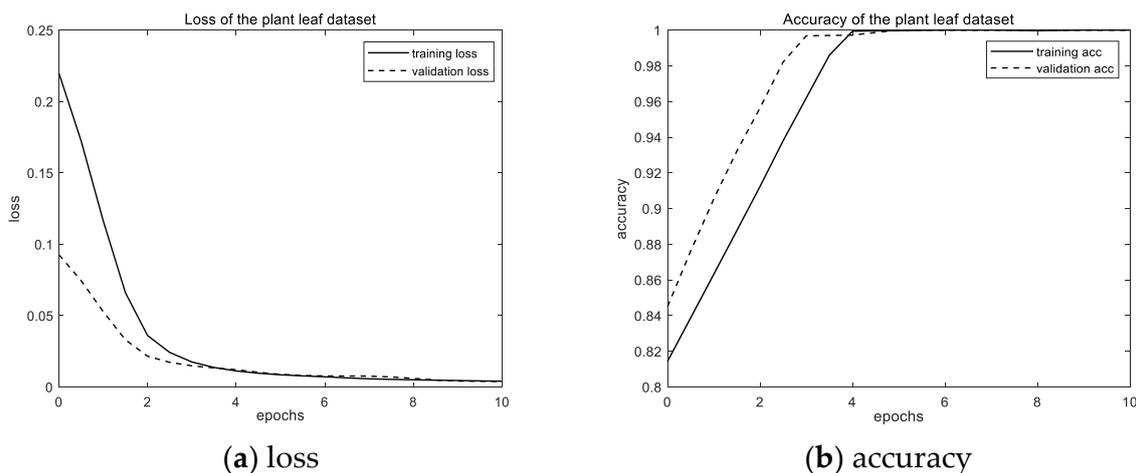


Figure 9. The distribution of embeddings from the crop pest dataset.

Through the distribution of the model’s output embeddings, it can be seen that the metric learning model has good ability for similarity matching on the single task, that is, the images from the same category gather while those from different categories are far away from each other.

Similar experiments on the other dataset were also carried out. The loss and accuracy of the CNN model on the plant leaf dataset is shown in Figure 10. The variation trends of the training loss and training accuracy are consistent with those of the validation loss and validation accuracy, which indicates that there is also no overfitting problem in the training period. The distribution of the model’s output embeddings of images from the plant leaf dataset is shown as Figure 11, which also shows the good ability of the similarity matching on a single task to distinguish the input paired images well.



(a) loss

(b) accuracy

Figure 10. The loss and accuracy on the plant leaf dataset.

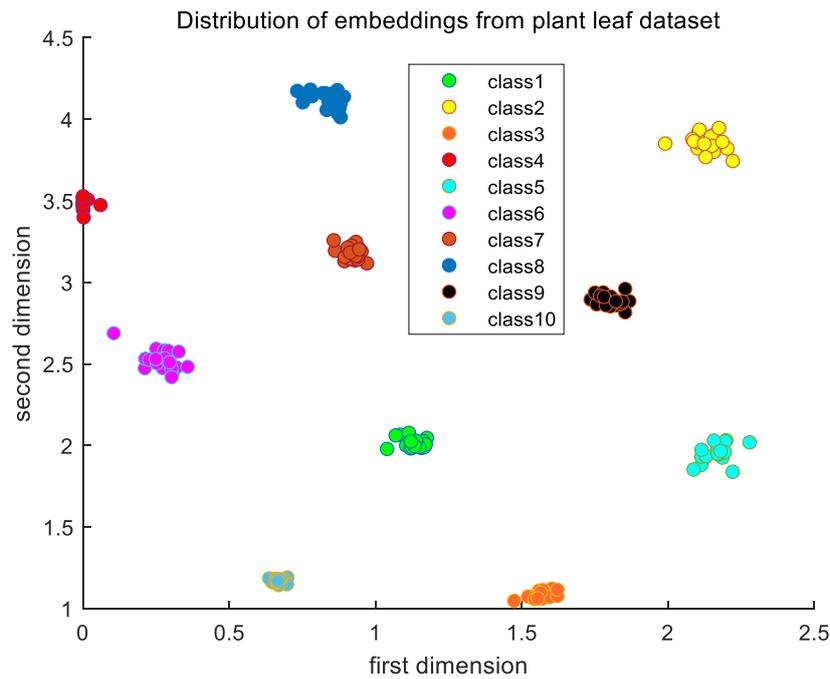


Figure 11. The distribution of embeddings from the plant leaf dataset.

3.2. Continual Tasks Experiment with the Basic CNN Model

As mentioned earlier, we hope that the model can be more flexible and able to handle continuous tasks, accumulating knowledge like humans to perform well on both old and new tasks. So, we carried out the experiments on sequential tasks to testify the continual performance of the CNN model, namely, the basic metric learning model.

For these two datasets, two occurring orders exist, that is, from the crop pest task to the plant leaf task, and the opposite one. For the first case, the testing accuracy of the two tasks is shown in Figure 12.

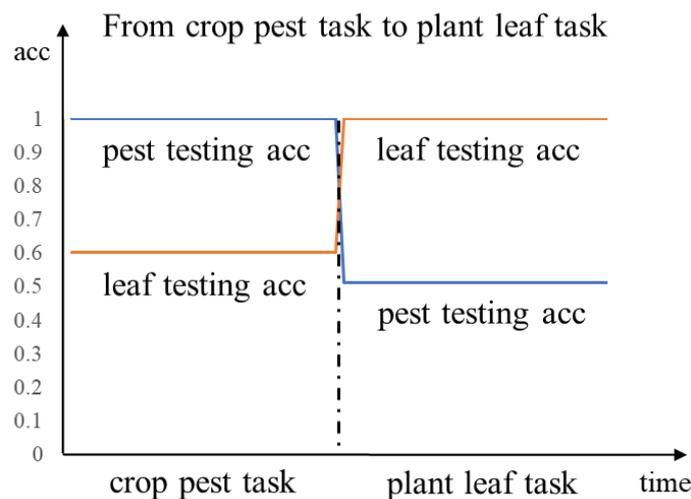


Figure 12. The testing accuracy of the first case.

At the first stage, the model has a good performance on the crop pest dataset, which was verified in Section 3.1. However, it has a very bad performance on the other dataset. The reason is that the other dataset is an unknown task and has never been seen before; this result is understandable and acceptable.

At the second stage, the model begins to learn the plant leaf task. Note that the model also learnt the crop pest task in the past. After the training period, the testing accuracy on the plant leaf task increases to 100% while that of the crop pest dataset decreases to nearly 50%, which is almost a blind guess. So, the extent of catastrophic forgetting for the crop pest task is nearly 50%. The new distribution of output embeddings from the old crop pest task is shown in Figure 13, which indicates that the basic metric learning model has lost the ability to distinguish the similarity between input paired images. The extracted features (embedding) of samples from different categories are mixed, and cannot be separated. This is an undesired forgetting problem!

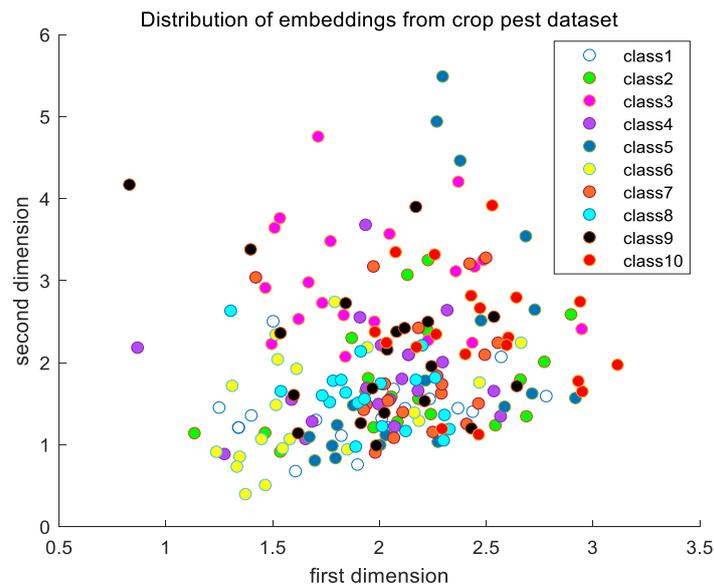


Figure 13. The distribution of embeddings from the old pest task.

For the second case, from the plant leaf task to the crop pest task, the experimental result of the testing accuracy is shown in Figure 14. The testing accuracy on the plant leaf task decreases from 100% to 60%, which means the extent of catastrophic forgetting for the plant leaf task is 40%. We found that, regardless of the occurring order of sequential tasks, the basic metric learning model does have a serious forgetting problem, as shown in Figures 12 and 14. In other words, after new learning, the basic metric learning model can no longer do the previous task well, due to the forgetting.

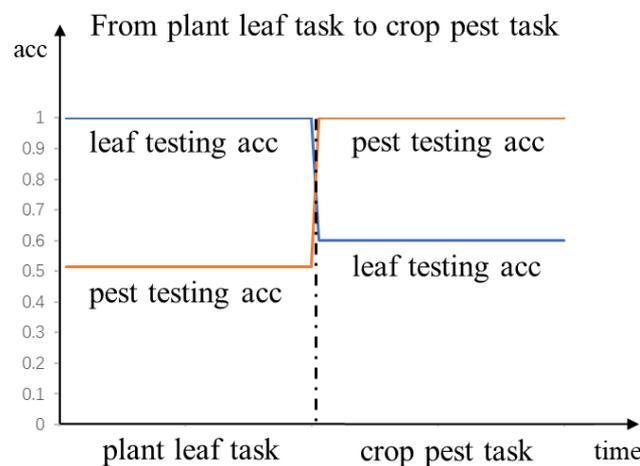


Figure 14. The testing accuracy of the second case.

The distribution of embeddings from the old plant leaf task is shown in Figure 15, which is very mixed and chaotic, losing the ability to distinguish and classify the similarity between input paired images.

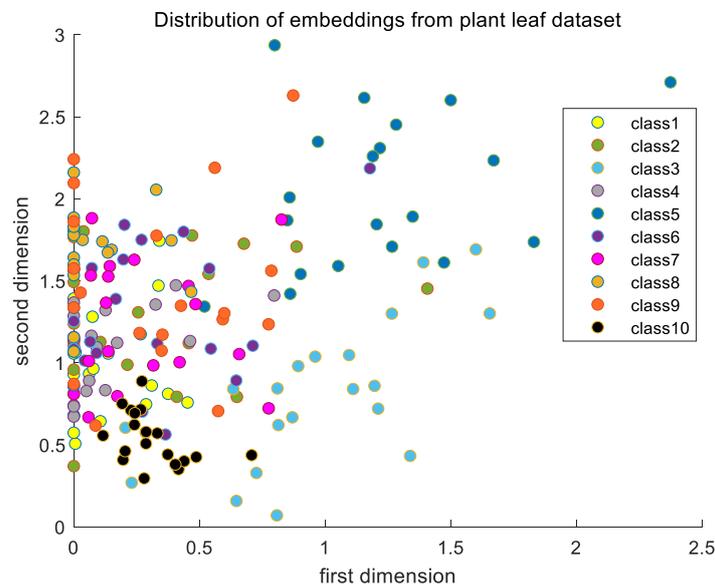


Figure 15. The distribution of embeddings from the old leaf task.

3.3. Continual Tasks Experiment with Our Proposed Method

As known, due to the forgetting problem, the basic CNN model cannot balance new and old tasks. Taking the sequential tasks from the crop pest dataset to the plant leaf dataset as an example, we used the designed GAN model to abstract the most important information of the old task (crop pest) and generated the abstracted images as memory for the future task, automatically ignoring the trivial details. When a new task comes, the abstracted images in memory will be retrieved and mixed with the new dataset, and then fed to the metric learning model.

Owing to this mechanism, the metric learning model can accumulate knowledge and better understand what it has learnt. The stored memory can be expanded, as does the increased ability to handle more continual tasks. The distribution of the model's output embeddings, corresponding to the testing images from both new and old tasks, is shown in Figure 16.

The results show the ability of our method to continually distinguish the similarity between input paired images and classify the testing images. All the categories from new and old tasks are separated clearly, which means that the metric learning model has a good performance on both new and old tasks, alleviating the forgetting problem. Compared with Section 3.2, the alleviated extent of catastrophic forgetting for the crop pest task and plant leaf task is 50% and 40%, respectively.

In addition, the results presented above are clear, and easily assessed. However, this is not always the case if we want to go further, e.g., an evaluation for the grouping results. In our opinion, the sum of the nearest distances between centers of groups will be a good choice. In detail, firstly, the center point of each group is calculated by the mean value; then, for every group center, the nearest distance with others is calculated; and finally, the sum of the nearest distances between the centers of groups is calculated, which is called the score. The evaluation metric should be proportional to the score, which means the larger the score is, the better the model's performance is.

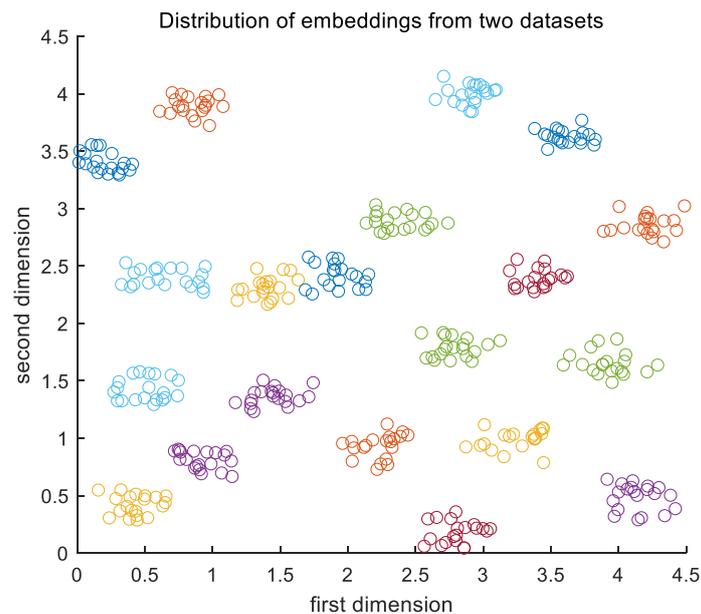


Figure 16. The distribution of embeddings from new and old tasks.

4. Discussion

We conduct the discussion about this work from the following three aspects.

4.1. Idea and Contents

The existing traditional models cannot accumulate the knowledge from old tasks, which means that they are all task specific, only focusing on the current task while forgetting the prior ones. This is a lack of flexibility and is quite different from humans' learning style. Besides, at present, there are mainly two basic types of neural network learning principles: Probability based on back-propagation error and similarity-based metric comparison. The former is more mature, but metric-based similarity learning is closer to biological learning.

So, from the bio-inspired perspective, we imitated the way biology learns and remembers, and proposed a continual metric learning method based on memory storage and retrieval to balance old and new tasks. Through several comparative experiments, it was found that the basic metric learning model can perform a single task excellently, distinguishing different categories well. However, when it is faced with continual tasks, the obvious forgetting problem occurs, and its poor flexibility loses the ability of dealing with old tasks. However, the addition of memory storage and retrieval in our method helps alleviate the forgetting problem, as all the categories from old and new tasks can be separated clearly, with good performance on both old and new tasks.

4.2. Contributions to Existing Research

We proposed an ANN-based continual classification method via memory storage and retrieval, combining the CNN and GAN technology, on the common agricultural datasets, such as the crop pest dataset and plant leaf dataset. The key contributions are two points: Few data and high flexibility.

As known, the big scale of the dataset is the basic requirement for the existing typical deep neural networks. However, the collection and labelling of big datasets are laborious and time consuming. So, research based on few data is a promising way. The metric learning used in this work only requires few raw data, because what it cares about is the paired inputs. Although the size of the raw dataset is small, the number of combinations of pairs from the same category and different categories can be expanded hundreds of times. Besides, the proposed continual learning method based on memory storage and retrieval increases the flexibility of the classification model, allowing it to balance old and

new tasks, by accumulating knowledge and alleviating forgetting. This can be regarded as another small step towards more intelligent and flexible studies in agriculture.

4.3. Limitations and Future Works

Although the numerical solution described in this study achieved a good performance on both new and old tasks with its continual learning ability, it still has some limitations. The key structure consisted of CNN and GAN. The CNN is relatively easy to implement with a stable performance, while the GAN is usually not stable and hard to train, and indeed requires experience. Besides, the number of sequential tasks is two, namely the crop pest dataset and plant leaf dataset, hence the work at this stage is only primary continual learning in agriculture. In future, we would like to analyze more tasks to develop the robust continual learning model, considering the complex combination of popular technologies in neural networks and information extraction.

5. Conclusions

In this study, we proposed an ANN-based continual classification method via memory storage and retrieval, combining the CNN and GAN, with two clear advantages. One is few data, as the metric learning model based on CNN works well from few data, which significantly reduces the difficulty of image collection and annotation; the other is flexibility, as continual classification based on memory storage and retrieval can balance old and new tasks through the accumulation of knowledge and alleviation of forgetting. The results show that the regular CNN can deal with a single task well and classify the categories clearly. However, when it comes to continuous tasks, there is a serious forgetting problem. With the addition of memory storage and the retrieval mechanism, the modified continual model can distinguish all the categories from both old and new tasks, without the forgetting problem. There are so many possible applications of this proposed approach in the field of agriculture, for instance, intelligent fruit picking robots, which can recognize and pick different kinds of fruits; and plant protection by the identification of diseases and pests, which can continuously improve the detection range. This work lays a foundation and provides a reference for other relevant studies towards more intelligent and flexible applications in the agricultural area.

Author Contributions: Conceptualization, Y.L.; Methodology, Y.L.; Writing—original draft, Y.L.; Software, X.C.; Validation, X.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Natural Science Program of Shihezi University, grant number KX01230101. The APC was funded by Shihezi University.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Martineau, M.; Conte, D.; Raveaux, R.; Arnault, I.; Munier, D.; Venturini, G. A survey on image-based insect classification. *Pattern Recogn.* **2017**, *65*, 273–284. [[CrossRef](#)]
2. Wang, J.; Lin, C.; Ji, L.; Liang, A. A new automatic identification system of insect images at the order level. *Knowl. Based Syst.* **2012**, *33*, 102–110. [[CrossRef](#)]
3. Ferentinos, K. Deep learning models for plant disease detection and diagnosis. *Comput. Electron. Agric.* **2018**, *145*, 311–318. [[CrossRef](#)]
4. Lu, Y.; Yi, S.; Zeng, N.; Liu, Y.; Zhang, Y. Identification of rice diseases using deep convolutional neural networks. *Neurocomputing* **2017**, *267*, 378–384. [[CrossRef](#)]
5. Ghazi, M.; Yanikoglu, B.; Aptoula, E. Plant identification using deep neural networks via optimization of transfer learning parameters. *Neurocomputing* **2017**, *235*, 228–235. [[CrossRef](#)]
6. Wu, Y.; Xu, L. Crop Organ Segmentation and Disease Identification Based on Weakly Supervised Deep Neural Network. *Agronomy* **2019**, *9*, 737. [[CrossRef](#)]
7. Pourdarbani, R.; Sabzi, S.; García-Amicis, V.M.; García-Mateos, G.; Molina-Martínez, J.M.; Ruiz-Canales, A. Automatic Classification of Chickpea Varieties Using Computer Vision Techniques. *Agronomy* **2019**, *9*, 672. [[CrossRef](#)]

8. Koirala, A.; Walsh, K.B.; Wang, Z.; Anderson, N. Deep Learning for Mango (*Mangifera indica*) Panicle Stage Classification. *Agronomy* **2020**, *10*, 143. [[CrossRef](#)]
9. Knoll, F.J.; Czymbek, V.; Harders, L.O.; Hussmann, S. Real-time classification of weeds in organic carrot production using deep learning algorithms. *Comput. Electron. Agric.* **2019**, *167*, 105097. [[CrossRef](#)]
10. Jwade, S.A.; Guzzomi, A.; Mian, A. On farm automatic sheep breed classification using deep learning. *Comput. Electron. Agric.* **2019**, *167*, 105055. [[CrossRef](#)]
11. Przybylak, A.; Kozłowski, R.; Osuch, E.; Osuch, A.; Rybacki, P.; Przygodziński, P. Quality Evaluation of Potato Tubers Using Neural Image Analysis Method. *Agriculture* **2020**, *10*, 112. [[CrossRef](#)]
12. Liu, Z.; Gao, J.; Yang, G.; Zhang, H.; He, Y. Localization and classification of paddy field pests using a saliency map and deep convolutional neural network. *Sci. Rep.* **2016**, *6*, 20410. [[CrossRef](#)] [[PubMed](#)]
13. Xie, C.; Wang, R.; Zhang, J.; Chen, P.; Dong, W.; Li, R.; Chen, H. Multi-level learning features for automatic classification of field crop pests. *Comput. Electron. Agric.* **2018**, *152*, 233–241. [[CrossRef](#)]
14. Abdalla, A.; Cen, H.; Wan, L.; Rashid, R.; Weng, H.; Zhou, W.; He, Y. Fine-tuning convolutional neural network with transfer learning for semantic segmentation of ground-level oilseed rape images in a field with high weed pressure. *Comput. Electron. Agric.* **2019**, *167*, 105091. [[CrossRef](#)]
15. Thenmozhi, K.; Reddy, U.S. Crop pest classification based on deep convolutional neural network and transfer learning. *Comput. Electron. Agric.* **2019**, *164*, 104906. [[CrossRef](#)]
16. Li, Y.; Yang, J. Few-shot cotton pest recognition and terminal realization. *Comput. Electron. Agric.* **2020**, *169*, 105240. [[CrossRef](#)]
17. Too, E.C.; Yujian, L.; Njuki, S.; Yingchun, L. A comparative study of fine-tuning deep learning models for plant disease identification. *Comput. Electron. Agric.* **2019**, *161*, 272–279. [[CrossRef](#)]
18. Kamilaris, A.; Prenafeta-Boldú, F.X. Deep learning in agriculture: A survey. *Comput. Electron. Agric.* **2018**, *147*, 70–90. [[CrossRef](#)]
19. Lima, M.C.F.; de Almeida Leandro, M.E.D.; Valero, C.; Coronel, L.C.P.; Bazzo, C.O.G. Automatic Detection and Monitoring of Insect Pests—A Review. *Agriculture* **2020**, *10*, 161. [[CrossRef](#)]
20. Fagot, J.; Cook, R.G. Evidence for large long-term memory capacities in baboons and pigeons and its implications for learning and the evolution of cognition. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 17564–17567. [[CrossRef](#)]
21. Grutzendler, J.; Kasthuri, N.; Gan, W.B. Long-term dendritic spine stability in the adult cortex. *Nature* **2002**, *420*, 812–816. [[CrossRef](#)] [[PubMed](#)]
22. Abraham, W.C.; Robins, A. Memory retention—The synaptic stability versus plasticity dilemma. *Trends Neurosci.* **2005**, *28*, 73–78. [[CrossRef](#)] [[PubMed](#)]
23. Carpenter, G.A. Looking to the future: Learning from experience, averting catastrophe. *Neural Netw.* **2019**, *120*, 5–8. [[CrossRef](#)]
24. Kaya, M.; Bilge, H.Ş. Deep metric learning: A survey. *Symmetry* **2019**, *11*, 1066. [[CrossRef](#)]
25. Parisi, G.I.; Kemker, R.; Part, J.L.; Kanan, C.; Wermter, S. Continual lifelong learning with neural networks: A review. *Neural Netw.* **2019**, *113*, 54–71. [[CrossRef](#)]
26. Li, Y.; Wang, H.; Dang, L.M.; Sadeghi-Niaraki, A.; Moon, H. Crop pest recognition in natural scenes using convolutional neural networks. *Comput. Electron. Agric.* **2020**, *169*, 105174. [[CrossRef](#)]
27. Berlemont, S.; Lefebvre, G.; Duffner, S.; Garcia, C. Class-balanced siamese neural networks. *Neurocomputing* **2018**, *273*, 47–56. [[CrossRef](#)]
28. Li, Y. Code and dataset for the ANN-based Continual Classification in Agriculture. *Zenodo* **2020**. [[CrossRef](#)]

