GAN based semi-supervised learning crop disease classifier

Wangzhihui Mei 2019124044 Chang Xu 20191xxxxx Zijia He 20191xxxxx Hongyi Huang 2019xxxxxx
CCNU-UOW JI

Abstract

Demo abstract

1 Introduction

As a big agricultural country, China suffer biological disasters such as disease and insect Tons annually. The crop diseases, insects, and grass diseases are not only diverse and widely distributed, but also have complicated disaster conditions. This directly affects the sustainable and stable development of agriculture and the growth of the rural economy. Practice has shown that the rational application of pesticides is an important measure to ensure a bumper harvest in agriculture, and the premise of rational application of pesticides is the correct diagnosis of the types and occurrence of pests and diseases. Not only is it impossible to obtain a good agricultural harvest, it will also cause a series of more serious problems. At present, the identification of pests and diseases still mainly rely on manual or semi-manual identification, which is actually a highly repetitive and time-consuming task. As an important subfield in artificial intelligence, machine learning plays a key role in modern intelligent technology. We can also use machine learning technology to detect crop diseases and insect pests, thereby saving manpower and achieving information agriculture.

For machine learning, according to the specific conditions of labeled and unlabeled samples contained in the training set, the task types can be roughly divided into the following types: supervised learning, unsupervised learning, and semi-supervised learning. In supervised learning, the classifier learns a large number of labeled training samples to build a model for predicting the label of unseen samples. "Label" refers to the output corresponding to the sample. In the classification problem, the label is the category to which the sample belongs, and in regression analysis, the label is the real-value output corresponding to the sample. With the rapid development of data collection and storage technology, it has been quite easy to collect a large number of unlabeled samples, and obtaining a large number of labeled samples is relatively difficult, because obtaining these labels may require a lot of manpower and material resources. There are also a large number of such cases in detection of crop diseases and insect pests. Relevant agencies have a large number of unlabeled data, and only a small number of representative samples can be labeled by human labor, but all labeling is not realistic. In fact, there are a lot of such situations in crop disease and pest

monitoring. The relevant agricultural management agencies have a large amount of unlabeled data, but if all these images are required to be labeled, the workload is extremely large and often unrealistic.

Semi-supervised learning (SSL) came into being under the above background and has become an important research field in pattern recognition and machine learning. In recent years, with the extensive application of machine learning theory in data analysis and data mining, such as web page and text classification, image and video retrieval, medical data processing, etc., semi-supervised learning has made great progress in theoretical research and practical applications . The research of semi-supervised learning mainly focuses on how to obtain a learner with good performance and generalization ability when part of the training data is missing. The lack of information here includes the absence of category labels or the presence of noise, and the lack of some feature dimensions of the data. The theoretical research of semi-supervised learning provides us with an in-depth understanding of many important theoretical issues of machine learning, such as the relationship between data manifolds and data categories, the proper handling of missing data, the effective use of labeled data, and the relationship between supervised and unsupervised learning Connection, the design of active learning algorithms, etc. have very important guiding significance.

2 Background theory

The research history of SSL can be traced back to the 1970s. During this period, Self-Training, Transdictionary Learning, Generative Mode and other learning methods appeared. The study of SSL became more fanatical in the 1990s, the emergence of new theories, and the development of new applications in natural language processing, text classification, and computer vision have promoted the development of SSL and emerged collaborative training New methods and Transient SupportSupperPortVechorMachine(TSVM) and other new methods. Merz et al. Proposed the term SSL in 1992 and used SSL for classification problems for the first time. Today, semi-supervised learning algorithms are mainly divided into 4 categories, which still has huge potential in the field of machine learning.

2.1 Generative Methods

Generative semi-supervised learning method is a method based directly on generative model. Generative semi-supervised learning(GSSL) assumes that all data (whether or not labeled) are generated by the same underlying model.

- This assumption makes it possible to associate unlabeled samples with learning objectives through the parameters of the latent model.
- The labeling of unlabeled samples can be regarded as the missing parameter of the model, which can usually be solved based on the maximum likelihood estimation based on the EM algorithm.

Models for generating examples include Gaussian models, Bayesian networks, Sigmoid BeliefNetWorks,

GMM, Multimedia Mixture Model(MMM), and Hidden Markov Model(HMM).

1. The examples in the Gaussian model follow the Gaussian distribution:
   $p(\boldsymbol{x}|y) = N(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{D/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right)$
2. The probability distribution of the samples in the Bayesian network is shown in the figure:

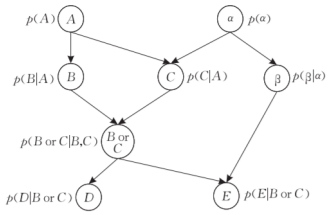

Figure 1: Bayesian Network

3. The samples in the S-type belief network obey the probability distribution:
   $p\left(\boldsymbol{x}_i|pa\left(\boldsymbol{x}_i\right)\right) = \frac{\exp\left(\left(\sum_j \boldsymbol{J}_{ij}\boldsymbol{x}_j + h_i\right)\boldsymbol{x}_i\right)}{1+\exp\left(\sum_j \boldsymbol{J}_{ij}\boldsymbol{x}_j + h_i\right)}$
4. GMM is a mixed distribution model of multiple Gaussian distributions, assuming that the sample is generated by weighted mixing of multiple models $\sum_i \pi_i p_i(\boldsymbol{x}|y)$, $\sum_i \pi_i = 1$.Each model follows a Gaussian distribution.
5. MMM is a mixed distribution model of multiple multi-modal distributions, assuming that the sample is generated by the weighted mixture of multiple models, The distribution of each model follows the multimodal distribution $p\left(\boldsymbol{x} = (x_{\cdot 1}, \cdots, x_{\cdot d})\,|\boldsymbol{\mu}\right) = \frac{\left(\sum_{i=1}^{D} x_{\cdot i}\right)!}{x_{\cdot 1}!\cdots x_{\cdot D}!}\prod_{d=1}^{D}\mu_d^{x_{\cdot d}}$
6. HMM is used to build a model of the sample sequence. The transition probability matrix between the specified states is transferred from one state to another state at a certain period to form the sequence. Each sample in the sequence is generated by the hidden state, where the state condition distribution can be Gaussian Mixed distribution or multi-modal mixed distribution. The current state depends only on the previous state, and the output only depends on the current state.

The advantages of generative semi-supervised learning method: the method is simple and easy to implement. In the case of very little labeled data, it often performs better than other methods. Disadvantages: The model assumption must be accurate, that is, the assumed generative model must be consistent with the real data distribution, otherwise the use of unlabeled data will reduce the generalization performance.

2.2 Discriminative Methods

The discriminative method uses the maximum interval algorithm to simultaneously train class targetsThe signed sample and the sample without class label learning decision boundary, as shown belowAs shown, make it pass through the low-density data area, and make the learnedThe distance between the classification hyperplane and the nearest sample is the largest.

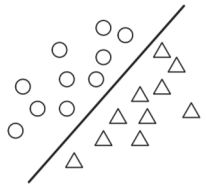

Figure 2: Schematic diagram of discriminant method

Discriminant methods include LDA, generalized discriminant analysis (GenerarizedDiscriminantAnalysis (GDA), semi-supervised supportVector machine (Semi-SuppliedSupperPortVecutorMachine,SVM), entropy regularization method and KNN method.

1. LDA is also called Fisher Linear Discrimination (Fisher Liner) Discriminative Analysis (FDA), originally developed by Fisher It was proposed in 1936 that the basic idea is to project the sample to the appropriate dimension In the low-dimensional space of, the projected sample has in the new subspace The largest inter-class distance and the smallest intra-class distance, that is, the sample according to the class Don't be divided into many clusters.
2. Badat and Anouar send LDA Expand to multiple types of problems, put forward GDA. Through a nonlinear mapping, the sample is mapped to a high-dimensional feature space and used in this feature space FDA training. When LDA and GDA are used for SSL, there is a the value of the class label of the divided sample is unknown, and the goal is to request the solution of the class value of the label.Because of $y_i \in \{c_1, c_2, \cdots, c_C\}$,this is a Mixed integer programming problem.
3. TSVM was originally proposed by Vapnik and Sterin for Linear prediction function for estimating class labels $f(x) = w^\top x + b$.Since it can also be used to estimate the class labels of unknown test samples, what is actually obtained is the decision boundary on the entire sample space. It is not a strict direct push method, but an inductive semi-supervised method, so it is called SVMs.TSVM attempts to consider various possible label assignments for unlabeled samples:
   - Trying to use each unlabeled sample as a positive or negative example.
   - Then, in all these results, we seek a partitioned hyperplane that maximizes the spacing between all samples (including labeled samples and unlabeled samples with label assignment).
   - Once the dividing hyperplane is determined, the final label assignment for unlabeled samples is its prediction result.

   Sample set of given markers $\mathbb{D}_l = \left\{\left(\overrightarrow{\mathbf{x}}_1, y_1\right), \left(\overrightarrow{\mathbf{x}}_2, y_2\right), \cdots, \left(\overrightarrow{\mathbf{x}}_l, y_l\right)\right\}$ and Unlabeled sample set $\mathbb{D}_u = \left\{\overrightarrow{\mathbf{x}}_{l+1}, \overrightarrow{\mathbf{x}}_{l+2}, \cdots, \overrightarrow{\mathbf{x}}_{l+u}\right\}$ among them $l \ll u, l + u = N, y_i \in \{-1, +1\}, i = 1, 2, \cdots, l$.The goal of TSVM learning is to give prediction labels for the samples in $\mathbb{D}_{\boldsymbol{u}}$.

$\overrightarrow{\hat{\mathbf{y}}} = (\hat{y}_{l+1}, \hat{y}_{l+2}, \cdots, \hat{y}_{l+u})^T$, $\hat{y}_i \in \{-1, +1\}, i = l+1, l+2, \cdots, N$ makes:

$$\begin{aligned} &\min_{\overrightarrow{\mathbf{w}}, b, \hat{\mathbf{y}}, \vec{\xi}} \frac{1}{2}\|\overrightarrow{\mathbf{w}}\|_2^2 + C_l \sum_{i=1}^{l} \xi_i + C_u \sum_{i=l+1}^{N} \xi_i \\ &s.t. y_i \left(\overrightarrow{\mathbf{w}}^T \overrightarrow{\mathbf{x}}_i + b\right) \geq 1 - \xi_i, \quad i = 1, 2, \cdots, l \\ &\hat{y}_i \left(\overrightarrow{\mathbf{w}}^T \overrightarrow{\mathbf{x}}_i + b\right) \geq 1 - \xi_i, \quad i = l+1, l+2, \cdots, N \\ &\xi_i \geq 0, \quad i = 1, 2, \cdots, N \end{aligned} \tag{1}$$

Among them:

- $(\vec{w}, b)$ defines a dividing hyperplane.
- $\vec{\xi}$ is the relaxation vector:
  - $\xi_i, i = 1, 2, \cdots, l$ corresponds to the marked sample.
  - $\xi_i, i = l+1, l+2, \cdots, N$ corresponds to the unlabeled sample.
- C_l, C_u are the compromise parameters specified by the user to balance the complexity of the model, the importance of marked samples, and the importance of unlabeled samples.

4. Entropy regularization method uses Shannon conditional entropy to measure the degree of overlap between classes, and its objective function is:

$$\max_\theta \frac{1}{2}\|w\|^2 + C_1 \sum_{i=1}^{l} \ln p\left(y_i|\boldsymbol{x}_i, \theta\right) + \lambda \sum_{i=l+1}^{n} \sum_{y_i=\tau_1}^{c_c} p\left(y_i|\boldsymbol{x}_i, \theta\right) \ln p\left(y_i|\boldsymbol{x}_i, \theta\right) \tag{2}$$

5. The KNN method finds the k nearest neighbor samples that are closest to the test sample in all samples, and the number of each type are $k_j, j = c_1, \cdots, c_C$,using decision rule $\arg\max_j k_j, j = c_1, \cdots, c_C$ to select the class label to mark the sample.

2.3 Graph-Based Methods

The essence of the graph-based method is label propagation (LabelPropagation). Based on the manifold hypothesis, the graph is constructed according to the geometric structure between the samples. The graph nodes are used to represent the samples. Propagate the class label from the sample with class label to the sample without class label.

The basic training process of the graph-based method is:

- Choosing an appropriate distance function to calculate the distance between samples. The available distance functions include Euclidean distance, Manhattan distance, Chebyshev distance, Ming's distance, Mahalanobis distance, and normalized Euclidean distance;
- According to the calculated distance, select an appropriate connection method, connect the samples with edges, and construct a connection diagram. The constructed connection graph is divided into dense graph (DenseGraph) and sparse graph (SparseGraph). The typical representative of dense graph is a fully connected graph, as shown in Figure 3, there are edge connections between any two nodes; sparse as shown in Figure 4, according to a certain criterion, the closest nodes are connected, including KNN diagram, ε-Nearest Neighbor (εNN) diagram, tangent weight diagram, exponential weight diagram, etc ;

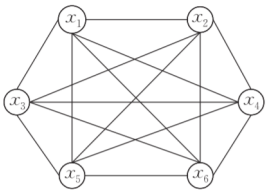

Figure 3: Full connection diagram

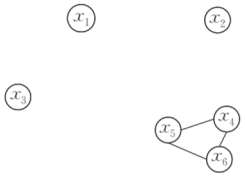

Figure 4: Sparse graph

- Kernel function is used to assign weights to the connected edges of the graph. The weights reflect the similarity between the two nodes connected by this edge. When the two nodes are $boldsymbolx_i$ And $boldsymbolx_j$ are very close, the weight w_{ij} of the edge connecting these two nodes is very large, these two examples have the same class label The probability is very large; conversely, when the two nodes are $boldsymbolx_i$ and $boldsymbolx_j$ are very far away, the edge connecting the two nodes The weight w_{ij} is very small, and the probability of these two samples having the same type of label is very small. Commonly used kernel functions include linear kernel, polynomial kernel, Gaussian kernel, radial basis kernel, hyperbolic tangent kernel, neural network kernel, Fisher kernel and spline kernel;
- Determine and solve the optimization problem according to the learning objective. The goal of the semi-supervised classification problem is to find the prediction function $f(x)$ of the class label that minimizes the objective function. This problem can be seen as a regularization risk minimization of a composite objective function consisting of a loss function and a regularization function Problem, the objective function for solving semi-supervised classification problems based on graph-based methods is generally expressed as:

$$\min_{f\,(x)} V(y, f(x)) + \lambda\Omega(f) \tag{3}$$

In equation (3), the loss function $V(y, f(x))$ is used to penalize the case where the predicted class label of the sample is not equal to the given class label. The smoothness of the function makes the predicted class labels of the nearest neighbors the same. According to the specific learning task, different loss functions and regularization functions can be selected. For example, the loss function can be the square error function, absolute value function, logarithmic function, exponential function, and hinge loss function. Generally, the loss function and the regularization function are limited to the re-copy kernel Hilbert space (Reproducing Kernel Hierarchy Spare, RHKHS), and the learning machine is

solved with a nuclear learning algorithm.

The following are some typical Graph-Based Methods:

1. In 2001, Blum and Chawla proposed the first graph-based SSL method-the minimum cut method (Mincut), which regards the samples marked as positive as the source node and the samples marked as negative as the target node Point, find a group of edges, so that after removing these edges, there is no connection between the source node and the target node, that is, the graph is divided into two independent clouds, and this group of edges is called a graph cut (Cut). After the graph is segmented, the label of the node class connected to the source node is marked as positive, and the label of the node class connected to the target node is marked as negative. This method selects the square loss with infinite weight as the loss function $V(y, f(\boldsymbol{x})) = \infty \cdot \sum_{i=1}^{l} (y_i - f(\boldsymbol{x}_i))^2$,represent regularization function with cut size $\Omega(f) = \boldsymbol{f}^{\mathrm{T}} L \boldsymbol{f}$ and limited to any node with class label x_i, the predicted class label is equal to the given class label $f(\boldsymbol{x}_i) = y_i$
2. In 2003, Zhu et al. Proposed the harmonic function method (Harmonic Function), also known as the Gaussian random field method (Gauss Random Field), to establish the distribution model of the prediction function on the graph, thereby expanding the discrete prediction function into a continuous prediction function. Like the minimum cut method, this method also chooses the square loss with infinite weight as the loss function. The size of the cut represents the regularization function. The difference is that the prediction function takes continuous values, rather than directly equals the class label value. The classification probability of the example solves the problem that the minimum cut method cannot solve.
3. Yan and Wang proposed a new SSL structure based on the l_1 graph. The idea of the l_1 graph comes from the fact that each data can be reconstructed by sparse linear superposition of training data, by The l_1 optimization problem obtains the sparse reconstruction coefficients, and uses the obtained coefficients to infer the weight of the directed l_1 graph, and simultaneously obtains the l_1 graph in a parameter-free manner. Near-neighbor structure and weights, and face recognition and image classification experiments, the implementation results show that the performance of this method is superior to the traditional graph method.

3 Application

There are two commonly used hypotheses in the field of semi-supervised learning: the cluster hypothesis and the manifold hypothesis. The clustering hypothesis refers to that the samples in the same cluster have a larger possibility of having the same mark, which means that the interface of different mark points should not appear in the area with a higher sample density while the manifold hypothesis means that adjacent samples have similar properties, and their labels should also be similar[5]. Since both the graph method and the density-based clustering method adhere to these two assumptions, the combined effect of the two methods on

a certain level should be comparable to the graph method alone. Specifically, the computational complexity of the graph method is higher, but the computational complexity of the local density estimation is much lower. Our picture uses the advantage of local giant thunder in computing time to reduce the composition of the node tree, on the other hand The method has a good classification effect on the overall characterization ability of the data structure.

According to this, we are completing unsupervised learning by labeling a small number of images and performing label propagation.

3.1 Preprocessing

Image preprocessing is an important link in the process of disease recognition. It is a kind of pre-processing relative to feature extraction and image recognition. As the actual production process, the collection equipment and environmental parameters are very different, there will be problems such as noise interference, low contrast, unclear targets, and interference from unrelated objects[4].

Pre-processing means that the chicken takes out the appropriate transformation of the image according to the actual situation to highlight some useful information and remove disturbing information. Therefore, we need to perform some tricky transformation to initial crop images.

3.1.1 Grayscale Conversion

Images are commonly saved by RGB(i.e. Red, Green and Blue) channels, which can preserve the natural color while reducing the memory consumption of the storage. Some researcher has done some related work[6], by extracting G-channel elements, they can not only simplify the algorithm, but also retain the original information of the image on the original basis. In this article, we also applied weighted sum conversion based on asymmetric weights. We increased the weight of G channel and reduce the weight of RB channels.

$$Gray_{img} = R \times 0.287 + G \times 0.599 + B \times 0.114$$

R,G,B represent RedGreen and Blue channel. The conversed image kept a lot of original information from initial image.

3.1.2 Image standardization and normalization

In the original crop image, the floating range of pixel values is large, which results in the objective function of the classification algorithm being unable to handle the features of the original image well in a large range. That is to say, assuming that a certain feature value in the original image has a large value range, such feature value will affect the final classification accuracy. Therefore, before classification, it is necessary to make a relatively standardized adjustment to his images, limiting the feature values to a certain

range, to maintain a relative balance between the original characteristics. In this article, we applied image normalization. X_mid is a middle variable to simplify the formula, X is pixel matrix, N is the total pixel num in the image, X_i is the value of pixel with the index i.

$$X_{mid} = \frac{X - \frac{1}{N}\sum_{i=1}^{N} x_i}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(x_i - \frac{1}{N}\sum_{i=1}^{N} x_i\right)^2}}$$

$$X_{norm} = \frac{X_{mid} - \min\left(X_{mid}\right)}{\max\left(X_{mid}\right) - \min\left(X_{mid}\right)} \times 255$$

3.2 Feature extraction

Crop diseases and insect pests will change the characteristics of the original plant tissue, including texture features and color features. The texture feature is a visual feature that does not depend on the color or brightness of the homogenous phenomenon in the image. It contains important information on the surface structure of the object and their relationship with the surrounding environment. Researchers can find other images with similar textures by submitting images containing certain textures. Color is the most prominent feature of people's sensory color images. Compared with other features, color features are relatively stable, are insensitive to image rotation, translation, and scale changes, and have good adaptability[3].

$$D_B = \frac{D_{max}}{A_0}\sum_{i=0}^{D_A} H_i \quad (D_A = 0, 1, 2, \ldots, L-1)$$

The conversion is simple, fast and effective, and has been widely used.

3.2.1 Color feature

Color is an important feature descriptor of image processing. When plants are attacked by diseases and insect pests, the abnormal areas of their images often change color. Construct. Based on this, we can construct histograms in different color spaces for feature extraction. After histogram equalization, the image may be insensitive to local differences in brightness, reducing the impact on classification results due to differences in image brightness. The pixels in the image after equalization occupy as many gray levels as possible and are evenly distributedso that the processed image will have a large contrast and dynamic range[1]. The formula of histogram equalization algorithm is as follows(D_B is the transformed gray value, D_max is the maximum gray value in the picture, A_0 is the total number of the pixels, D_A is the gray value before transformation, H_i is the number of pixels with ith gray value, L is the maximum pixel level in the image).

3.2.2 Texture feature

Texture is a visual feature that reflects the homogeneous phenomenon in the image, and texture analysis is the extraction and analysis of the gray scale spatial distribution pattern of the image. Texture features are

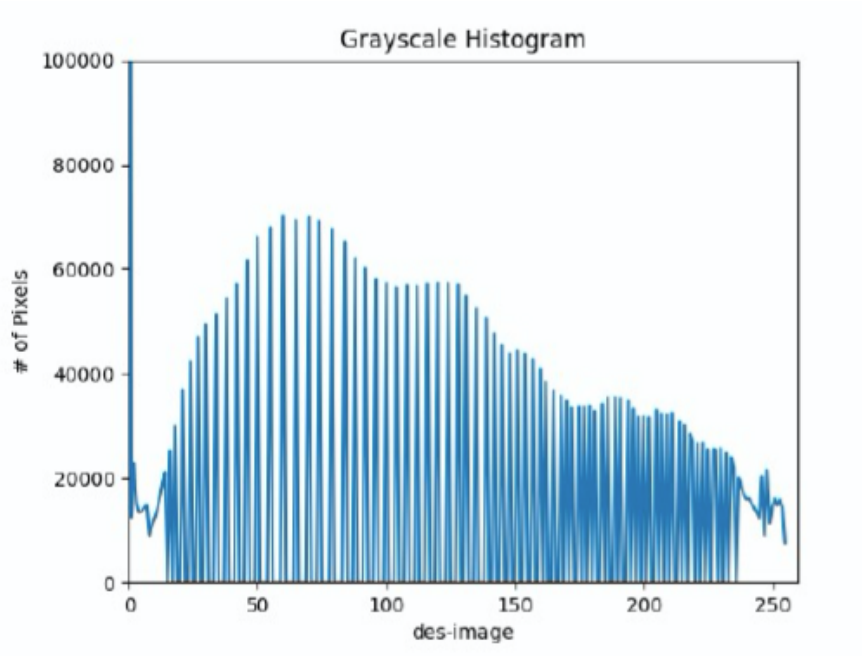

Figure 5: The grayscale histogram

different from image features such as gray scale and color[2]. It is expressed by the gray scale distribution of pixels and their surrounding spatial neighborhood. The texture feature embodies the property of the surface structure and arrangement of the surface structure that changes slowly or periodicatwo.

we extract the following two texture features from each crop image:

1. Gray level co-occurrence matrix: statistical texture features based on gray level co-occurrence matrix (GLCM). This feature is based on the gray attributes of pixels and their neighborhoods, and the statistical features in the texture area are studied. The gray level co-occurrence matrix is defined by the joint probability density of the pixels at two positions, and can reflect the comprehensive information of the image gray level about the direction, adjacent interval, and change range. In this paper, the gray level co-occurrence matrix of $\theta = 0°, 45°, 90°, 135°$ is first extracted, and then the energy, contrast, correlation, entropy and inverse gap are five typical parameters that can intuitively reflect the texture

condition described by the co-occurrence matrix, and 20 feature values are extracted from an image.

$$ASM = \sum_i \sum_j P(i,j)^2 \tag{4}$$

$$CON = \sum_i \sum_j (i-j)^2 P(i,j) \tag{5}$$

$$COR = \sum_i \sum_j \frac{(i-\mu_i)\left(j-\mu_j\right)}{\sigma_i \sigma_j} P(i,j) \tag{6}$$

$$ENT = -\sum_i \sum_j P(i,j) \log P(i,j) \tag{7}$$

$$IDM = \sum_i \sum_j \frac{P(i,j)}{1+(i-j)^2} \tag{8}$$

2. Gray gradient co-occurrence matrix: The gray gradient co-occurrence matrix model mainly reflects the relationship between the two most basic elements in the image, namely the gray level and the gradient (or edge) of the image point. The gray level of each image point is the basis for forming the image, and the gradient is the element that forms the edge contour of the image. The main information of the image is provided by the edge contour of the image. The gray gradient space clearly describes the resolution law of the gray and gradient of each pixel in the image, and also gives the spatial relationship between each image point and its field image point. It can describe the image texture well and reflect the direction texture from the gradient direction. The classification result of the gray-level gradient co-occurrence matrix is superior to the gray-level co-occurrence matrix, because the gray-level co-occurrence matrix uses only gray-scale information, while the gray-scale gradient co-occurrence matrix uses both the gray-scale and gradient information of the image. Gray gradient co-occurrence matrix (GGCM) texture feature analysis is to extract the texture features of gray gradient co-occurrence matrix. The gradient information of the image is added to the gray level co-occurrence matrix, so that the co-occurrence matrix can better contain the elements of the texture image and their arrangement information.

4 Conclusion

In this article, this paper uses semi-supervised generative adversarial networks to perform semi-supervised learning on crop disease and insect images. We redesigned the network composition of generators and discriminators in generative adversarial networks based on the characteristics of crop images. In the generative confrontation network used, this paper designs a multi-layer generator network and a discriminator deep neural network. The input to generate the adversarial network is the result of the improved network segmentation, combining the artificially selected features with the features calculated by the neural network.

5 References

References

[1] **Guo, Peng, Ronald J Stanley, Justin G Cole, Jason R Hagerty, and William V Stoecker**, "Color Feature-based Pillbox Image Color Recognition.," in "VISIGRAPP (4: VISAPP)" 2017, pp. 188–194.

[2] **Humeau-Heurtier, Anne**, "Texture feature extraction methods: A survey," *IEEE Access*, 2019, *7*, 8975–9000.

[3] **Mohanty, Sharada P, David P Hughes, and Marcel Salathé**, "Using deep learning for image-based plant disease detection," *Frontiers in plant science*, 2016, *7*, 1419.

[4] **Rechcigl, Jack E**, *Environmentally safe approaches to crop disease control*, CRC Press, 2018.

[5] **Shen, Xipeng, Matthew Boutell, Jiebo Luo, and Christopher Brown**, "Multilabel machine learning and its application to semantic scene classification," in "Storage and Retrieval Methods and Applications for Multimedia 2004," Vol. 5307 International Society for Optics and Photonics 2003, pp. 188–199.

[6] **Sun, Bo, Abdullah M Iliyasu, Fei Yan, Fangyan Dong, and Kaoru Hirota**, "An RGB multi-channel representation for images on quantum computers," *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 2013, *17* (3), 404–417.