

Analysis of Image Based Classification Using Machine Learning Techniques

Ashish Chaudhary ,Akash Kumar, Ankit

Guide name: Ms. Barkha Bhardwaj

Noida Institute of Engineering and Technology, Greater Noida.

Abstract:

This research study about image classification by using the deep neural network (DNN) or also known as Deep Learning by using framework TensorFlow. Python is used as a programming language because it comes together with TensorFlow framework. The input data mainly focuses in Fruit category which there are different types of Fruit that have been used in this paper. Deep neural network (DNN) has been choosing as the best option for the training process because it produced a high percentage of accuracy. Results are discussed in terms of the accuracy of the image classification in percentage. apple get 90.585% and same goes to another type of fruit where the average of the result is up to 90% and above. Image classification serves as the foundation for many computer vision tasks, enabling machines to interpret and understand visual data. By automatically categorizing images into predefined classes or labels, machine learning models facilitate efficient decision-making processes and enable automation in various real-world scenarios. Applications range from medical image diagnosis and surveillance systems to content recommendation algorithms and image search engines. Machine learning approaches to image classification typically involve the use of convolutional neural networks (CNNs), which have demonstrated remarkable performance in extracting relevant features from images and learning complex patterns. CNNs leverage hierarchical layers of convolutional filters to extract increasingly abstract representations of images, followed by fully connected layers for classification. Transfer learning, data augmentation, and ensemble methods are commonly employed techniques to enhance the performance and robustness of image classification models. The future of image classification using machine learning holds promise for advancements in various areas.

1. Introduction

Classification is a systematic arrangement in groups and categories based on its features. Image classification came into existence for decreasing the gap between the computer vision and human vision by training the computer with the data. The image classification is achieved by differentiating the image into the prescribed category based on the content of the vision. Motivation by [1], in this paper, we explore the study of image classification using deep learning. The conventional methods used for image classifying is part and piece of the field of artificial intelligence (AI) formally called as machine learning. The machine learning consists of feature extraction module that extracts the important features such as edges, textures etc. and a classification module that classify based on the feature extracted. The main limitation of machine learning is, while separating, it can only extract certain set of features on images and unable to extract differentiating features from the training set of data. This disadvantage is rectified by using the deep learning [2]. Deep learning (DL) is a sub field to the machine learning, capable of learning through its own method of computing. A deep learning model is introduced to persistently break down information with a homogeneous structure like how a human would make determinations. To accomplish this, deep learning utilizes a layered structure of several algorithms expressed as an artificial neural system (ANN). The architecture of an ANN is simulated with the help of the biological neural network of the human brain. This makes the deep learning most capable than the standard machine learning models [3]. In deep learning, we consider the neural networks that identify the image based on its features. This is accomplished for the building of a complete feature extraction model which is

capable of solving the difficulties faced due to the conventional methods. The extractor of the integrated model should be able to learn extracting the differentiating features from the training set of images accurately. Many methods like GIST, histogram of gradient oriented and Local Binary Patterns, Scale-Invariant Feature Transform (SIFT) are used to classify the feature descriptors from the image. In this paper, deep neural network, based on TensorFlow is used with Python as the programming language for image classification. Thousands of images are used as the input data in this project. The accuracy of each percentage of 'train' session will be studied and compared. Image classification has become a major challenge in machine vision and has a long history with it. The challenge includes a broad intra-class range of images caused by color, size, environmental conditions, and shape.

Machine learning is an important branch of artificial intelligence. Although machine learning has experienced half a century of development, there are still some unsolved problems, for example, complex image understanding and recognition, natural language translation, and recommendation system [4]. Deep learning is an important branch developed on the basis of machine learning. It makes full use of the hierarchical characteristics of artificial neural network and biological neural system to process information and obtains high-level features by learning low-level features and adopting feature combination method, so as to realize image classification or regression.

Related Works

Image classification provides an important basis for image depth processing and the application of computer vision technology in related fields. Traditional image classification mainly goes through different stages, such as image preprocessing, feature extraction, classifier construction, and learning training [5]. Traditional image classification methods mainly use the extracted basic image features to realize image classification, which can provide a basis for further obtaining

the semantic information of images by computer. Traditional image classification generally uses image color, texture, and other information to calculate image features and uses support vector machine and logistic regression to realize image classification [6]. The results of image classification not only depend on the extracted features to a great extent but also are affected by the knowledge and experience of relevant fields.

Not only are the manually acquired features difficult to apply to image classification but also a lot of time is spent in analyzing feature data. At the same time, the traditional machine learning cannot be applied to the processing of large datasets, and it is difficult to realize the optimization of feature design, feature selection, and model training, which makes the classification effect of the model poor. Therefore, image classification methods using traditional machine learning are affected in many application fields. Research shows that because texture, shape, and color features can be used for image classification and recognition, low-level basic features can be used as the basis of image classification. Traditional image classification methods generally use single feature extraction or feature combination and take the extracted features as the input value of support vector machine. In recent years, some progress has been made in image classification using artificial neural network classifier. In order to improve the accuracy of image classification, we can focus on the standardized design of low-level features such as texture, shape, and color [7]. Deep learning realizes the training of large-scale datasets through multilevel network model and adopts the method of layer-by-layer feature extraction to obtain the high-level features of the image. Not only is the deep learning network model used to extract the basic features of the image but also it can obtain the deep features of the image through multiple hidden layers. Compared with traditional machine learning methods, the features obtained by deep learning method are not only accurate but also conducive to image classification. In the process of image recognition and classification, the way of feature learning and combination is mainly determined by the deep learning model. At present, the commonly used deep learning models are sparse model, restricted Boltzmann machine model, and convolution neural network model. Although these models have some differences in feature extraction, they have similarities in image classification and recognition. They all go through the steps of image information input, data preprocessing, feature extraction, model training, and classification output.

In image classification, some scholars have carried out a lot of research work on image feature representation and classifier selection. For example, the deep learning model based on feature representation can be effectively applied to the recognition and classification of various images. Some scholars use deep convolution neural networks (DCN) to deeply extract image features and apply them to large-scale dataset ImageNet. Experiments show that the model can effectively classify large data image sets. In addition, the deep learning model can effectively learn and describe image features. For example, the deep learning model can better describe the hierarchical features through unsupervised learning, and the features extracted by the model not only have strong expression ability but also improve the efficiency of image classification [8].

Fundamentals of Image Classification Algorithm

3.1. Basic Theory of Neural Network. The traditional neural network, referred to as artificial neural network (ANN), is a hot spot in the field of early artificial intelligence. Artificial neural network mainly uses the neurons of network model to abstract the characteristics of external things, so as to be used by computer to complete information processing.

Artificial neural network generally establishes the corresponding network structure according to the different construction methods of neurons. Neural network is an operation model composed of several different nodes or neurons connected with each other. Each node in the model is a processing function, and the connection between different nodes uses weight to represent the memory ability of artificial neural network. The output of neural network depends on the connection form, weight value, and excitation function of different nodes. At the same time, the neural network model is mainly constructed according to some algorithm or function to express some specific logical operation.

A basic neural network model usually includes information input layer, hidden layer, and calculation result output layer. Different layers can contain several neurons. Neurons represent a transformation or operation, which is completed by the activation function of neurons. Two adjacent layers of neurons are connected to each other, as shown in Figure 1.

As can be seen from Figure 1, the neural network model includes 11 neurons: 3 input layers, 5 hidden layers, and 3 output layers. The structure belongs to a two-layer neural network model, where W_1 and W_2 are the weight matrices of the hidden layer and the output layer, respectively

[8]. Deep learning method is a part of machine learning. It is widely used in natural language recognition and image detection and classification. Moreover, deep learning comes from the theory of artificial neural network. By referring to the human brain for hierarchical processing of information, different levels of neural networks are established. Deep learning effectively extracts multilevel feature information by simulating human brain, so as to obtain the key feature information of image, text, and other data. Deep learning mainly describes the specific object characteristics through hierarchical processing according to a large amount of edge feature information. It is a process from low-level feature extraction to high-level feature combination. As an important method of machine learning, deep learning is an effective method to process big data and obtain abstract features by using neural network model.

Deep learning is a multilayer deep neural network model. According to the connection law of human brain neurons, the sample features are processed in different layers of the model, and the deep features of the sample data are obtained in turn. Similar to deep neural network, artificial neural network belongs to hierarchical structure. The model is composed of multilayer perceptron, including input layer, hidden layer, and output layer. Different from the deep neural network, the artificial neural network model only contains two to three layers of forward neural network, and the number of neurons in each layer is small, so the processing ability of large datasets is limited. Because the deep neural network model contains many layers and each layer contains a large number of neurons, the neural network model not only can realize the abstract expression of data but also has strong learning function. Compared with the traditional machine learning methods, the deep learning model does not need to rely on manual design and feature

Hidden layer

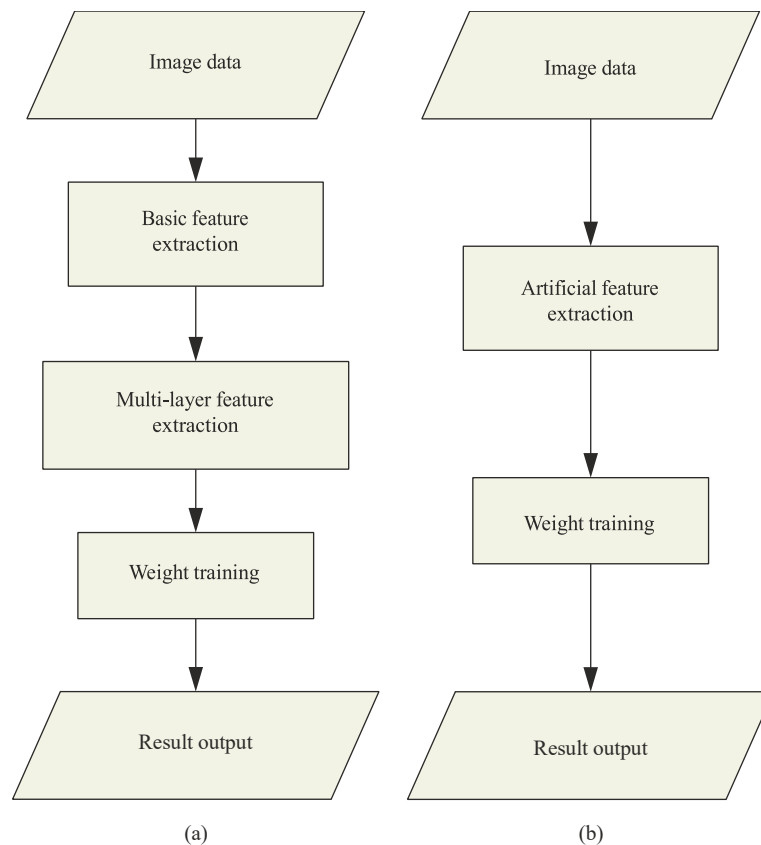


Figure 2: Comparison between deep learning model and traditional machine learning algorithm. (a) Deep learning algorithm.

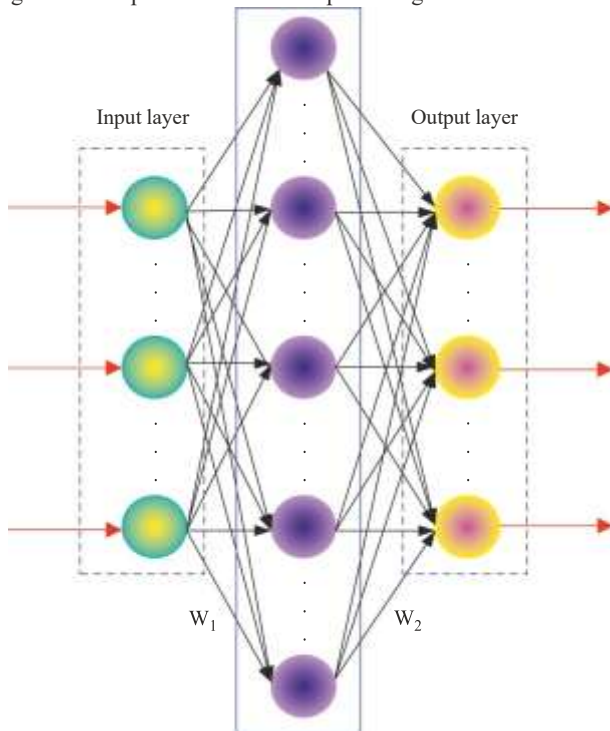


Figure 1: Connections between neurons in different layers.

extraction. It not only solves the shortage of manual or knowledge but also avoids the preference problem in feature extraction. At the same time, the deep learning model can obtain representative high-level features through the organic combination of low-level features [9]. As shown in Figure 2, the working processes of deep learning and traditional machine learning are compared.

In the neural network model, the activation function is used to perform nonlinear operation on the input data of neurons in

extract effective feature information from the original input data. Activation function is a nonlinear function. With the increase of the number of layers of neural network model, the most effective feature information can be obtained after many iterations and data training.

In order to realize the classification of features, Softmax function is often used as the activation function in the neural network model and used in the output layer of the model. Softmax function is generally used as a classifier. The calculation formula is as follows:

$$S_k = eR_k / \sum_{j=1}^n eR_j$$

In the above formula, n is the number of neurons in the current layer and R_k is the nonlinear transformation value of the k -th neuron in this layer.

Softmax is a classifier that can output different feature categories. The value of each neuron contained in Softmax can be considered as the probability of the corresponding category. The calculation process of Softmax function is shown in Figure 3.

(b) Traditional machine learning algorithm.

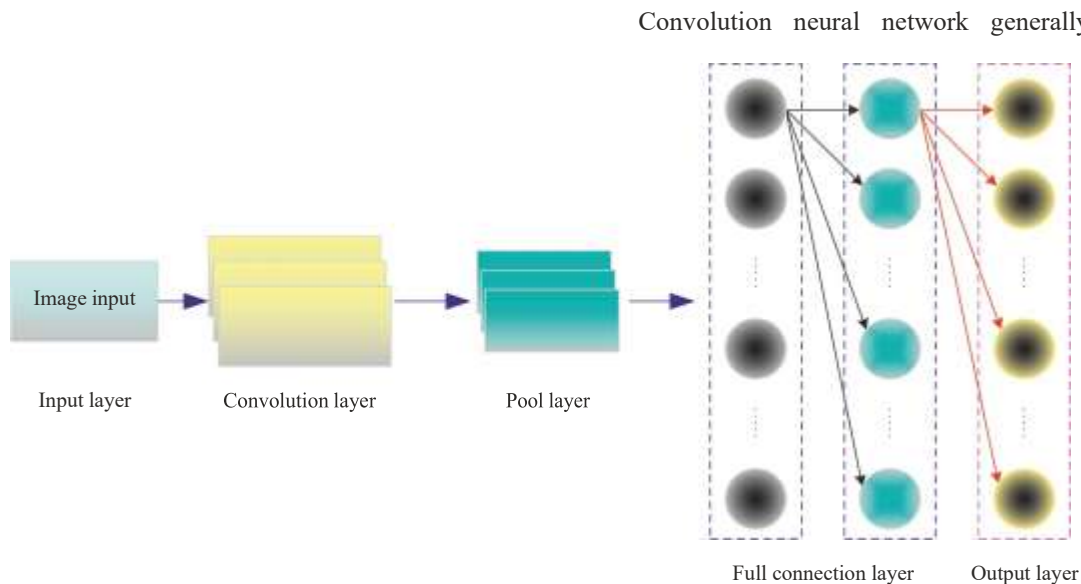


Figure 4: Structure diagram of convolution neural network.

Image (3×3)

a ₁	a ₂	a ₃
b ₁	b ₂	b ₃
c ₁	c ₂	c ₃

Convolution kernel (2×2)

Figure 5:
diagram of
operation

1	3
5	7

Feature output (2×2)

Schematic
convolution
process.

$a_1+3a_2+5b_1+7b_2$	$a_2+3a_3+5b_2+7b_3$
$b_1+3b_2+5c_1+7c_2$	$b_2+3b_3+5c_2+7c_3$

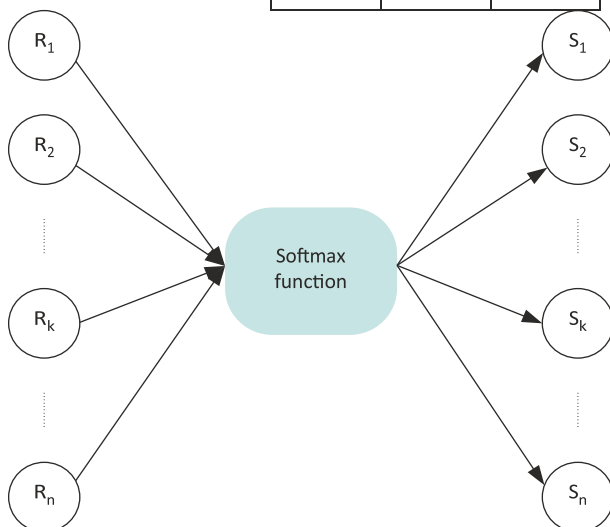


Figure 3: Schematic diagram of Softmax function calculation.

3.2. Basic Theory of Convolution Neural Network. Convolution neural network (CNN) is a typical network structure in deep learning model [10]. Different from traditional machine learning, convolution neural network can be better used for image and time series data

processing, especially for image classification and language The basic structure of convolution neural network is shown in Figure 4.

Convolution neural network generally includes three

different types of processing layers: convolution layer, pooling layer, and full connection layer. Among them, the feature extraction task is completed by the convolution layer, and the pooling layer is used for feature mapping. The full connection layer is similar to the general neural network structure. All nodes in this layer are not connected to each other but completely connected to the nodes of the previous layer. In addition, like other neural networks, convolution neural networks also have data input layer and result output layer.

The calculation task of convolution neural network is mainly completed through the convolution layer, and the convolution kernel in the convolution layer is the core of the convolution neural network model. The convolution layer uses convolution check to convolute the input image and extract the characteristic information of the image. The images processed by convolution operation will gradually become smaller, and the pixels at the edge of the image have little effect on the output results.

As shown in Figure 5, assuming that the original input image is a 3×3 matrix, the original image is convoluted through a convolution check with a size of 2×2 , and the corresponding feature map is output.

Generally, there is a strong correlation between adjacent pixels in the image. Convolution kernel mainly extracts features from the local area of the image and sends the extracted local features to the high level for integration processing. Because the bottom feature of the image is independent of its position, it can not only use the same convolution check to extract the relevant features but also reduce the number of parameters of the neural network through the shared weight characteristic of the

convolution kernel, so as to improve the training efficiency of the network model.

For complex images, in order to reduce the amount of parameter training of the model, the pooling layer in convolution neural network can be used to reduce the size of feature map. During pooling, the depth and size of the image can remain unchanged. The operation of pooling layer generally includes max pooling and average pooling [11], as shown in Figure 6.

3.3. Convolution Neural Network Model. In the existing image detection and recognition models, such as ResNet, Mask-RCNN, and Faster R-CNN models, they are usually based on common network models [12]. LeNet is the most basic convolution neural network model [9]. After transforming the LeNet model, the LeNet-5 model is established to classify ordinary images. Because the LeNet-5 model is not deep enough, it cannot extract enough image features during model

training. Therefore, it cannot be applied to the classification of complex images.

Therefore, some people put forward the AlexNet model based on the LeNet structure, applied the convolution neural network to the processing of complex images, and provided a theoretical basis for the application of deep learning model in the field of computer vision [17]. The network structure of AlexNet is shown in Figure 7.

AlexNet is a network structure with 8 layers. The model includes 5 convolution layers and 3 full connection layers. The model uses the ReLU function as the activation function to avoid the gradient dispersion phenomenon caused by the large number of layers of the network model. In order to reduce network training time, AlexNet uses multiple GPUs for training. In order to suppress neurons with small response ability, AlexNet uses LRN (local response normalization) processing layer to establish a competition mechanism for neurons, so as to make the value with large response ability increase continuously, so as to increase the generalization ability of the model. In addition, in order to prevent neurons from forward propagation and back

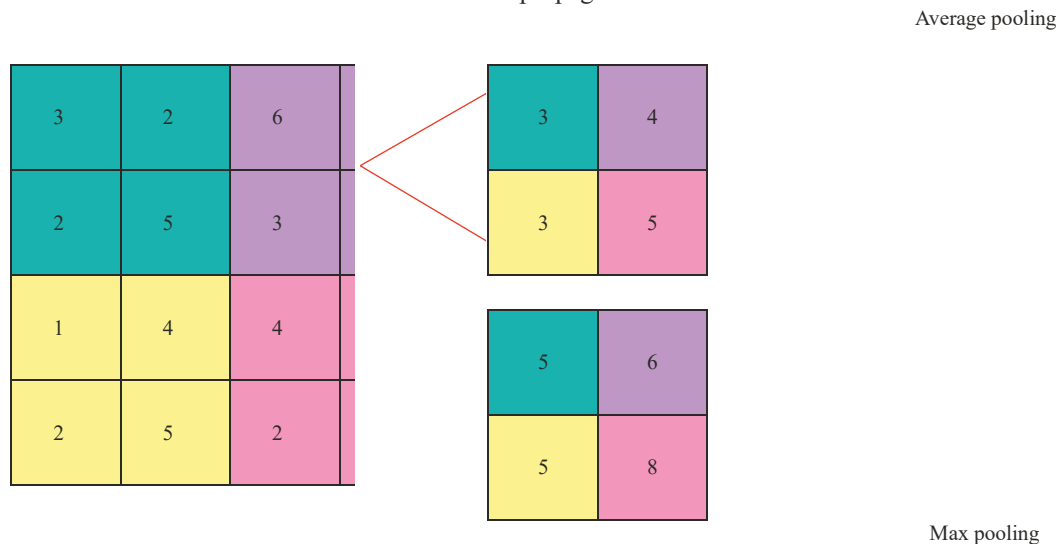


Figure 6: Schematic diagram of pool layer operation type.

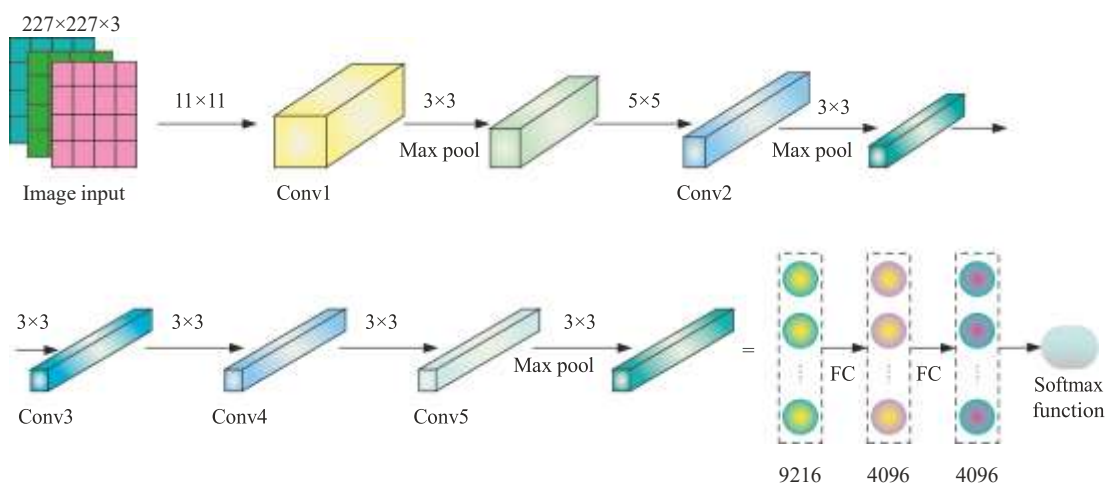


Figure 7: Schematic diagram of AlexNet network model structure.

propagation, the dropout mechanism is adopted in the full

Table 1: The network structure of Vgg-16.

connection layer of AlexNet model, so that the output results of hidden layer neurons are 0, so as to avoid the complex interaction between neurons.

At present, VggNet is a widely used deep convolution neural network model. Compared with other models, VggNet not only better generalization ability but also can be effectively used for recognition of different types of images [13]. For example, convolution neural networks such as FCN, UNet, and SegNet are on VggNet model. In recent years, Vgg-16 and Vgg-19 networks been commonly used for VggNet models [18]. The network structure of Vgg-16 is shown in Table 1.

The Vgg-16 network structure has 16 layers in total, excluding the pooling layer and Softmax layer.

The convolution core size is 3×3 , the pooling layer size is 2×2 , and the pooling layer adopts the maximum pooling operation with step size of 2.

Vgg-16 network uses convolution blocks instead of convolution layers, in which each convolution block contains 2~3 convolution layers, which is conducive to reducing the network model parameters. At the same time, Vgg-16 network adopts ReLU activation function to enhance the training ability of the model. Although Vgg model has more layers than AlexNet model, the convolution kernel of Vgg model is smaller than that of AlexNet model. Therefore, the number of training iterations of Vgg model is less than that of AlexNet model.

2. Image Classification Model Based on Improved CNN

4.1. Improved Image Classification Model Framework. Because image classification algorithms are usually used in systems with high real-time requirements, image classification algorithms need to consider real-time performance. For complex neural network models, image classification needs to consume a lot of time. Therefore, this paper simplifies VggNet model and takes it as the model basis of image classification.

Considering the distribution characteristics of datasets used for model training, a typical dataset can be selected as the weight of the model to initialize the training dataset. When the model is pretrained and reaches a certain accuracy, the number of nodes in Softmax layer is reduced by ten times, and then the dataset is used for weight training. Considering that the data processed by the model may be affected by various noises, a noise reduction automatic encoder is added to the model to eliminate the noise interference, and the existing dataset is extended through the data enhancement method to enhance the generalization ability of the model.

Considering that the image classification algorithm needs to meet certain real-time performance, the corresponding image classification model is established and optimized based on VggNet model. Among them, the algorithm combining convolution neural network and noise reduction automatic encoder can be used. Because there may be overfitting problem in image classification, it can be optimized by data enhancement. Compared with other algorithms, this classification algorithm has certain generalization performance in the case of small amount of data. In addition, the algorithm also adds a noise reduction automatic encoder, which can effectively reduce the impact of data noise on the performance of the model, so as to ensure that the model has good generalization ability. Since the improved algorithm is based on VggNet network, the training time of the model may increase

Layer type	Convolution kernel size	Characteristic diagram size	Convolution kernel number	
Input layer	—	448×448	—	all
Convolution layer	3×3	448×448	128	has
Pool layer	3×3	448×448	—	the
Convolution layer	3×3	224×224	256	based
Pool layer	2×2	112×112	—	have
Convolution layer	3×3	112×112	768	
Pool layer	2×2	56×56	—	
Full connection layer	4096	—	—	
Full connection layer	4096	—	—	
Full connection layer	1000	—	—	[14].
Softmax classifier	—	—	—	The

improved image classification algorithm is shown in Figure 8.

In order to solve the problem of automatic noise reduction of complex image structure, the normalized network of encoder is used for classification in this paper. Based on the existing convolution neural network model, the noise reduction automatic encoder and sparse automatic encoder are organically combined, and the input original image information is normalized on the sparse automatic encoder. Then, the improved convolution neural network model is used to extract the image feature information, and the Softmax classifier is used to classify the features. When the improved convolution neural network model is used to classify images, it is very necessary to preprocess the image such as noise reduction and grayscale, select a certain number of training sets and test sets from the dataset, and then take the training set as the input object of the model after unsupervised learning processing. Secondly, the hidden layer of the noise reduction automatic encoder is used to encode and decode the input object, and the processing results are output to the sparse automatic encoder of the next layer for normalization. The data is trained layer by layer through the hidden layer of sparse automatic encoder, and finally the training results of sparse automatic encoder are output to Softmax classifier. In order to improve the classification accuracy, gradient descent method can be used to strengthen the training of classifier model parameters in order to improve the performance of image classification depth learning model. Finally, the network model is verified by using the image test set, and the effectiveness of the image classification method is tested according to the classification results output by the model. The improved convolution neural network model can overcome the problem that the traditional neural network is only limited to some features in image classification. Through the normalization of sparse automatic encoder, the overfitting phenomenon of model in data processing can be avoided, and more abstract representative features can be obtained by using the hidden layer of sparse automatic encoder to train the data layer by layer. The improved model adopts Softmax classifier, which can make the classification result closer to the real value. The improved deep learning network model is mainly divided into two stages: training and testing. The training stage is mainly used to build an effective image classification model, and the testing stage is mainly to evaluate and analyze the model according to the experimental classification results. Figure 9 shows the workflow of the improved deep learning network model.

4.2. Image Classification Model Optimization. It is known from the existing research that the convolution neural network model can be optimized from the aspects of data enhancement and adjusting training methods, and the optimization of convolution neural network model is related to the type of model used. For example, for the deep learning model with more layers, the training parameters can be optimized.

According to the structure of convolution neural network, convolution layer is used to extract features, and the size of convolution kernel determines the extraction quality of image features. From the perspective of image composition, adjacent pixels can generally form the edge lines of the image. Several edge lines form the image texture, and the image texture is combined into several local patterns. The local pattern is the basic element of the image. Through the convolution layer of the network model, different types of features can be extracted and the local pattern of the image can be formed. When the convolution kernel is smaller, although the convolution layer

the optimization of convolution layers can improve the classification accuracy of the model.

In order to improve the accuracy of image classification and recognition, the depth learning model proposed in this paper needs to be optimized. Firstly, a smaller convolution kernel is selected in the first convolution layer in order to extract more image feature details. Secondly, the maximum pool sampling operation is adopted in the model to avoid the overfitting problem. As shown in Figure 10, the optimized image classification model consists of three convolution layers, in which the convolution kernel of each convolution layer decreases in turn. At the same time, after each convolution layer, the features are processed by ReLU activation function, and the generated features are used as the input of the maximum pooling layer. The model adopts three full connection layers, takes the processing result of the last full connection layer as the input of Softmax classifier, and then generates the classification result of the image.

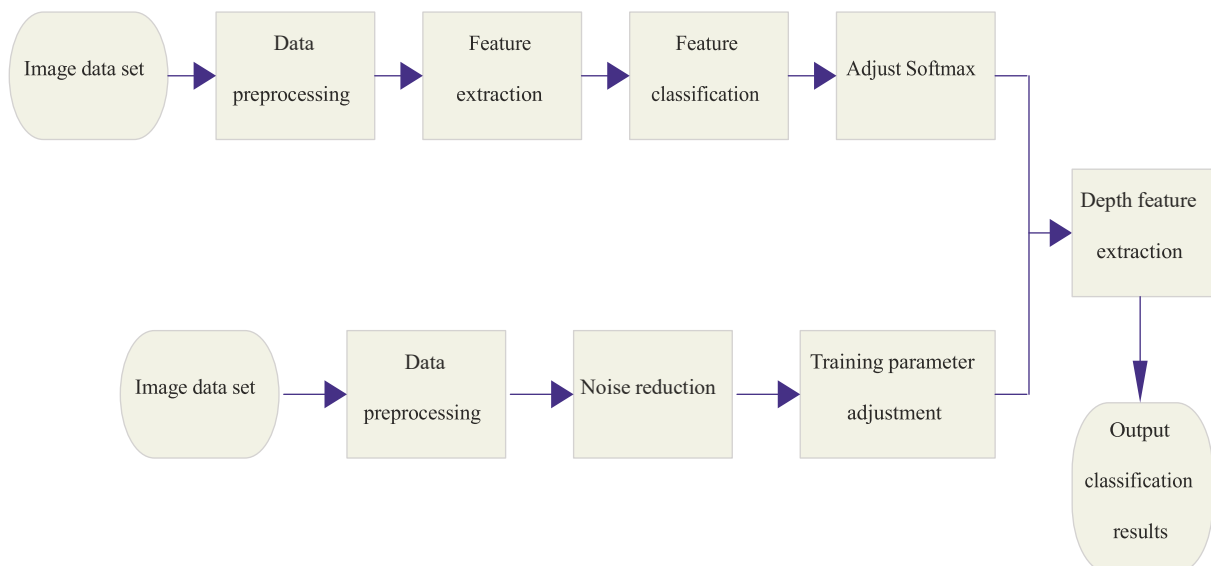


Figure 8: Working diagram of improved image classification algorithm.

extracts more features, there may be overfitting problems in the model. On the contrary, the larger the convolution kernel is, the fewer features extracted by the convolution layer are, and the worse the image classification effect may be. Therefore, the reasonable optimization of the size of convolution kernel can improve the accuracy of image classification.

Because the convolution neural network model mainly extracts the image features layer by layer through different convolution layers, the number of convolution layers will affect the feature extraction quality of the model to a certain extent. Similar to the number of convolution kernels contained in the convolution layer, the more convolution layers, the finer the features obtained by the model classifier, which may lead to overfitting phenomenon, while the less convolution layers, the coarser the features obtained by the model classifier, which may lead to the decline of image classification accuracy. Therefore,

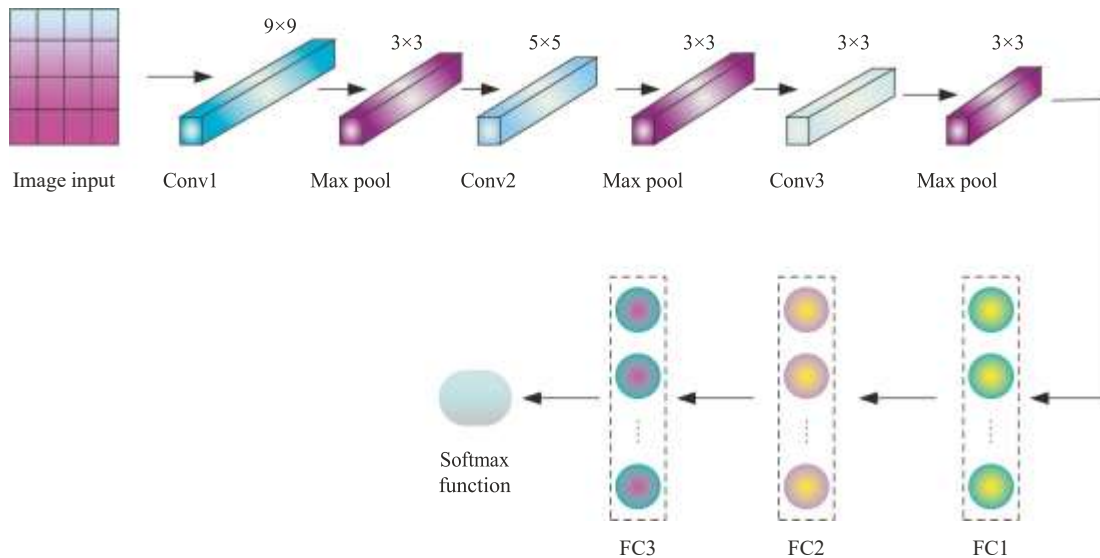


Figure 9: Structure diagram of optimized image classification model.

[15]. The convolution layer parameters of the model include the size and number of convolution kernels. The first convolution layer is close to the image input layer and is mainly used to extract the basic features of the image, so the parameters of the first convolution layer have a great influence on the features. In order to facilitate the further processing of features in the subsequent convolution layer, a smaller convolution kernel needs to be used to extract the attribute information such as shadow, boundary, and light of the image.

The convolution layer maps the obtained features through the activation function. Therefore, the optimized convolution neural network model adopts ReLU activation function [20], which can be expressed by mathematical function as follows:

$$y(x) = \max(0, x). \quad (2)$$

When the traditional convolution neural network model uses ReLU activation function to train features, it may lose useful feature information in the process of image classification. In order to prevent the loss of useful features during image classification, it can be improved on the basis of the existing ReLU activation function. The optimized activation function can be expressed as

$$y(x) = \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases} \quad (3)$$

From the improved calculation formula of activation function, when the input feature is less than zero, it can not only retain the negative value information in the feature map but also increase the reinforcement learning of effective features.

The optimized convolution neural model uses Softmax function to classify the images, and Softmax function uses supervised learning algorithm to regress the features. In the classification process, category y of the image target can have M different values. If the image training set is $\{(x_1, y_1), \dots, (x_i, y_i)\}$, where x_i represents the image training sample, y_i is the image classification category, and $y_i \in \{1, 2, \dots, M\}$, the cost function of Softmax regression algorithm can be expressed

Assuming that M markers are accumulated in the cost function, the probability calculation of training sample x as category k can be obtained, which is expressed as

$$p(y = k | x; \alpha) = \frac{\exp(\alpha_k x_k)}{\sum_{j=1}^M \exp(\alpha_j x_j)} \quad (5)$$

3. Experiment and Analysis

5.1. Selection of Dataset and Experimental Methods. To verify the effectiveness of the deep learning-based image classification model proposed in this paper, the **Fruit Image Dataset** was selected for experimentation. This dataset contains various images of different fruit types captured under varying conditions such as lighting, background, shape, and angle, introducing challenges like intra-class variance and inter-class similarity.

The dataset used contains 50+ categories of fruits, with 80 images per category, totaling 5625 images. Each image is labeled with the corresponding fruit class (e.g., Apple, Banana, Cherry, Grape, Kiwi, Lemon, Mango, Orange, Papaya, Pear, Pineapple, Pomegranate, Strawberry, Watermelon, etc.). These images were randomly divided into three subsets: training set (70%), validation set (15%), and test set (15%).

The experiments were conducted using Python and the TensorFlow/Keras deep learning framework. Key steps included:

- Image preprocessing: resizing to 224×224, normalization, and augmentation (rotation, flipping, zoom)
- Feature extraction: using convolutional layers
- Model training and evaluation

Figure 10 shows a sample of the fruit images used in the dataset.

Figure 10: Partial images of the Fruit Dataset used for classification (e.g., Apple, Mango, corn, Banana, Strawberry, etc.)



To evaluate the model's classification robustness, accuracy was used as the main metric. We computed:

- **Overall Classification Accuracy (AC):** Ratio of correctly classified samples to the total number of samples.
- **Per-Class Classification Accuracy (CCAi):** Ratio of correctly classified samples of type i to total test samples of type i .

The formulas are as follows:

$$AC = \frac{t_n}{tr}, CCAi = \frac{t_{ri}}{t_{ni}}$$

In the above formula, t_{ni} represents the number of correctly classified samples, t_n denotes the number of test samples, t_{ri} indicates the number of correctly classified samples of type i , and t_{ni} expresses the number of test samples of type i . [16]

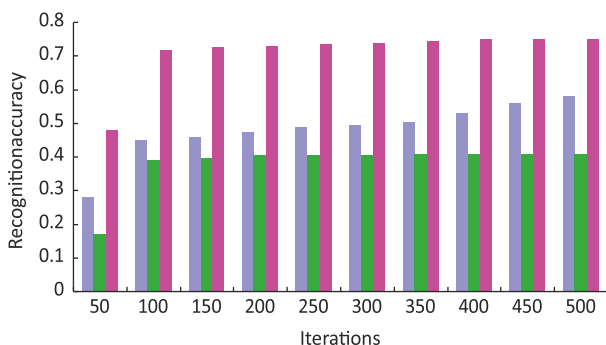


Figure 12: The relationship between the classification accuracy of the model and the number of iterations.

5.2. Results and Analysis.

The proposed model was coded in Python and tested in Jupyter Notebook on the fruits dataset which contains 3670 images from 20 fruit classes. The model could achieve an accuracy of 91.84% for 6 fruit classes. The neural network could successfully identify the fruit image and classified it to the correct fruit class with an accuracy of 99.89%. The model successfully detected 1845 testing images from 6 fruit classes, 3076 training images from 6 fruit classes, and 2549 validated images from 6 fruit classes.

12. When iterating 50 times, the accuracy of LeNet model is 28%, that of AlexNet model is 17%, and that of VggNet model is 48%. When the number of iterations is 100, the accuracy of LeNet model is 45%, that of AlexNet model is 39%, and that of VggNet model is 72%. In addition, when the VggNet model converges, the highest accuracy rate is as high as 75%, the highest accuracy rate of LeNet model is 58%, and the highest accuracy rate of AlexNet model is 41%.

In order to test the effect of image classification after the model is optimized to a certain extent, the accuracy of flower image classification before and after the model optimization is compared in the experiment, as shown in Figure 13. The comparison results show that, for the training dataset, the optimized model converges faster in the early stage of training and slower in the middle stage of training, while the two models in the later stage of training are basically the same. For the test dataset, the optimized model is higher than the nonoptimized model in terms of convergence speed and image classification accuracy. Therefore, the model optimization method proposed in this paper can effectively improve the accuracy of image classification.

In addition, in order to test the relationship between the loss value function of the optimized model and the number of iterations, the training set and test set are used to compare the models before and after optimization, as shown in Figure 14. The loss value function of the nonoptimized model shows an upward trend with the increase of the number of iterations, indicating that the nonoptimized model has the phenomenon of overfitting, while the loss value function of the optimized model shows a downward trend with the increase of the number of iterations. It can be seen that the cost of parameter training can be reduced through model optimization.

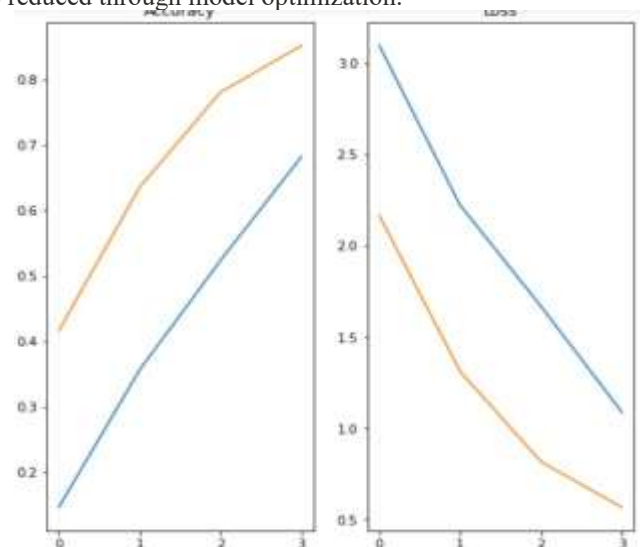
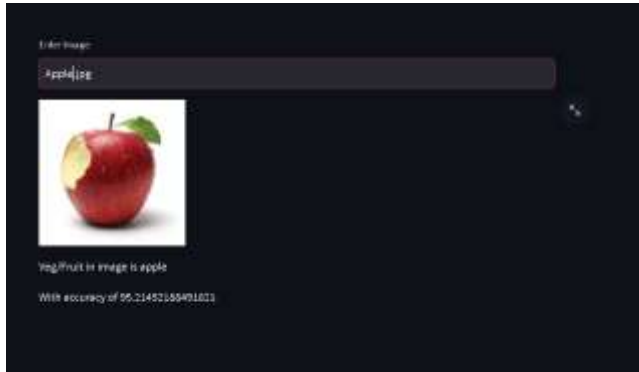


Figure 12: Comparison of accuracy and loss.

From the above experimental comparison results of the relationship between the accuracy of common network models in image classification and the number of iterations, it is known that the model proposed in this paper is superior to other models in classification accuracy. By comparing the classification accuracy of the deep learning model on the training set and the test set before and after optimization, it is known that the accuracy of image classification can be significantly improved after a certain degree of optimization. [17]



significantly influenced the direction and outcomes of this project.

- Members of our research team for their dedication, hard work, and collaboration in collecting and annotating the image dataset, designing experiments, and analyzing results.
- Contributors to open-source libraries, frameworks, and tools used in this project, including TensorFlow, PyTorch, scikit-learn, and Matplotlib, for their continuous development and support.

4. Conclusion

In conclusion, this research is about image classification by using deep learning via framework TensorFlow. It has three (3) objectives that have achieved throughout this research. The objectives are linked directly with conclusions because it can determine whether all objectives are successfully achieved or not. It can be concluded that all results that have been obtained, showed quite impressive outcomes. The deep neural network (DNN) becomes the main agenda for this research, especially in image classification technology. DNN technique was studied in more details starting from assembling, training model and to classify images into categories. The roles of epochs in DNN was able to control accuracy and also prevent any problems such as overfitting. Implementation of deep learning by using framework TensorFlow also gave good results as it is able to simulate, train and classified with up to 90% percent of accuracy towards two different types of Animals that have become a trained model. Lastly, Python have been used as the programming language throughout this research since it comes together with framework TensorFlow which leads to designing of the system involved Python from start until ends.

Data Availability

The labeled dataset used to support the findings of this study is available from the corresponding author upon request.

Future scope

- Image can be classified and keep in separate folder.
- Automatic face recognition and Object recognition can be used for classifying the images automatically.
- It is useful for classifying larger number of images within short time.

Acknowledgments

- Our project mentor for their guidance, encouragement, and invaluable feedback, which

References

- [1] P. Rastegari, M., Ordonez, V., Redmon, J., & Farhadi, A. (2016). XNOR-net: Imagenet classification using binary convolutional neural networks. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9908 LNCS, 525–542. https://doi.org/10.1007/978-3-319-46493-0_32
- [2] Gupta, K. K., Vijay, R., & Pahadiya, P. (2022). Detection of abnormality in breast thermograms using Canny edge detection algorithm for thermography images. *International Journal of Medical Engineering and Informatics*, 14(1), 31-42.
- [3] Pasolli, E., Melgani, F., Tuia, D., Pacifici, F., & Emery, W. J. (2014). SVM active learning approach for image classification using spatial information. *IEEE Transactions on Geoscience and Remote Sensing*, 52(4), 2217–2223. <https://doi.org/10.1109/TGRS.2013.2258676>
- [4] Korytkowski, M., Rutkowski, L., & Scherer, R. (2016). Fast image classification by boosting fuzzy classifiers.
- [5] Gupta, K. K., Vijay, R., Pahadiya, P., Saxena, S., & Gupta, M. (2023). Novel Feature Selection Using Machine Learning Algorithm for Breast Cancer Screening of Thermography Images. *Wireless Personal Communications*, 1-28.

<https://doi.org/10.1007/s11277-023-10527-9>.

- [6] Dumoulin, V., & Visin, F. (2016). A guide to convolution arithmetic for deep learning, 1–31. <https://doi.org/10.1051/0004-6361/201527329>
- [7] Gupta, K. K., Pahadiya, P., & Saxena, S. (2022). Detection of cancer in breast thermograms using mathematical threshold-based segmentation and morphology technique. *International Journal of System Assurance Engineering and Management*, 13(1), 421-428. <https://doi.org/10.1007/s13198-021-01289-3>.
- [8] Kui, X., Liu, W., Xia, J., & Du, H. (2017). Research on the improvement of python language programming course teaching methods based on visualization. *ICCSE 2017 - 12th International Conference on Computer Science and Education*, (Iccse), 639–644. <https://doi.org/10.1109/ICCSE.2017.8085571>
- [9] Laila Ma` rifatul Azizah, Sitti Fadillah Umayah, Slamet Riyadi, Cahya Damarjati, Nafi Ananda Utama “Deep Learning Implementation using Convolutional Neural Network in Mangosteen Surface Defect Detection”, ICCSCE, ISBN 978-1-5386-3898-9, pp. 242-246, 2017
- [10] Gupta, K. K., Vijay, R., & Pahadiya, P. (2020). A Review Paper on Feature Selection Techniques and Artificial Neural Networks Architectures Used in Thermography for Early-Stage Detection of Breast Cancer. *Soft Computing: Theories and Applications*, 455-465, (Scopus Index).
- [11] Rika Sustika, Asri R. Yuliani, Efendi Zaenudin, Hilman F. Pardede “On Comparison of Deep Learning Architectures for Distant Speech Recognition”, 2nd International Conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE), ISBN 978-1-5386-0659-9, pp. 17- 21, 2017
- [12] Gregor, K., Danihelka, I., Graves, A., Rezende, D. J., & Wierstra, D. (2015). DRAW: A Recurrent Neural

Network For Image Generation.

<https://doi.org/10.1038/nature14236>

- [13] Rui Wang, Wei Li, Runnan Qin and JinZhong Wu “Blur Image Classification based on Deep Learning”, IEEE, ISBN 978-1-5386-1621-5 pp. 1-6, 2017.
- [14] Aizat Faiz Ramli, Hafiz Basarudin, Mohd Azlan Abu, Muhyi Yaakop, Mohamad Ismail Sulaiman, FUSA: Fuzzy Logic Based Clustering Protocol for Formation of Uniform Size Clusters, 2017 International Conference on Engineering Technology and Technopreneurship (ICE2T), pp 1-6, 2017.
- [15] Siregar, S. T. M., Syahputra, M. F., & Rahmat, R. F. (2018). Human face recognition using eigenface in cloud computing environment. *IOP Conference Series: Materials Science and Engineering*, 308(1), 012013. <https://doi.org/10.1088/1757899X/308/1/012013>
- [16] Sharma, M., Anuradha, J., Manne, H. K., & Kashyap, G. S. C. (2017). Facial detection using deep learning. *IOP Conference Series: Materials Science and Engineering*, 263, 042092 <https://doi.org/10.1088/1757-899X/263/4/042092>
- [17] Gupta, K.K., Vijay, R., Pahadiya, P. et al. Use of Novel Thermography Features of Extraction and Different Artificial Neural Network Algorithms in Breast Cancer Screening. *Wireless Personal Communication* 123, 495–524 (2022). <https://doi.org/10.1007/s11277-021-09141-4>