



OPEN ACCESS

EDITED BY

Francesca Borghi,
University of Milan, Italy

REVIEWED BY

Gianluca Milano,
National Institute of Metrological Research
(INRIM), Italy

Fabio Michieletti,
Polytechnic University of Turin, Italy, in
collaboration with reviewer GM
Massimiliano Galluzzi,
Chinese Academy of Sciences (CAS), China

*CORRESPONDENCE

Xiangbing Gong,
✉ xbgong@csust.edu.cn

RECEIVED 11 May 2024

ACCEPTED 23 September 2024

PUBLISHED 14 October 2024

CITATION

Xu P, Wang J, Jiang Y and Gong X (2024)
Applications of artificial intelligence and
machine learning in image processing.
Front. Mater. 11:1431179.
doi: 10.3389/fmats.2024.1431179

COPYRIGHT

© 2024 Xu, Wang, Jiang and Gong. This is an
open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with
these terms.

Applications of artificial intelligence and machine learning in image processing

Pingyuan Xu¹, Jinyuan Wang², Yu Jiang² and Xiangbing Gong^{2*}

¹Zhejiang Infrastructure Construction Group Co., Ltd., Hangzhou, China, ²School of Traffic and Transportation Engineering, Changsha University of Science and Technology, Changsha, China

Introduction: With the rapid development of artificial intelligence and machine learning technology, image processing technology based on artificial intelligence and machine learning has been applied in various fields, which effectively solves the multi-classification problem of similar targets in traditional image processing technology.

Methods: This paper summarizes the various algorithms of artificial intelligence and machine learning in image processing, the development process of neural network model, the principle of model and the advantages and disadvantages of different algorithms, and introduces the specific application of image processing technology based on these algorithms in different scientific research fields.

Results And Discussion: The application of artificial intelligence and machine learning in image processing is summarized and prospected, in order to provide some reference for researchers who used artificial intelligence and machine learning for image processing in different fields.

KEYWORDS

artificial intelligence and machine learning, artificial intelligence algorithm, neural network model, image processing and application, deep learning

1 Introduction

With the development of information technology, great progress has been made in image feature analysis, image registration, image fusion, image classification, image recognition, content-based image retrieval, and image digital watermarking. These image processing techniques reflect human intelligent activity and mimic, extend and extend human intelligence on computers. They have intelligent processing capabilities. Due to the remarkable achievements of artificial intelligence research and application at home and abroad, the application effect of artificial intelligence technology in image processing will be huge.

The idea of artificial intelligence and machine learning first emerged in the 1940s and 1950s, when people proposed the idea of artificial intelligence, machine learning also emerged. Many related scholars hoped to make computers intelligent and capable of continuous self-learning like humans. Turing (1950) created the Turing test to determine whether a computer is intelligent. But at that time, it was only in the initial stage. The level of artificial intelligence and machine learning was very low, and it was very difficult to realize intelligence. However, in the 1980s, Rumelhart et al. proposed the error back propagation algorithm (Rumelhart et al., 1986), which brought a new direction to machine

learning. Until now, we still use this algorithm to realize machine learning. This algorithm does not have the predetermined rules of human beings, and the machine can learn independently from a large number of data and calculate the laws according to relevant methods, from which to find out the natural laws, and then used to predict and decide the future data, completing the independent learning of the machine in the real sense. But this kind of learning is superficial. In the 1990s, Pontil and Verri (1998) proposed support vector machine algorithm (Pontil and Verri, 1998); Schapire (1999) proposed Adaboost algorithm (Schapire, 1999). Machine learning theory research has achieved fruitful results. Many shallow models have been proposed. Support vector machine, SVM, LR and other maximum entropy methods have all achieved great success, and good feedback has been obtained in theory and practice. But these shallow learning models require a lot of practice and rich experience to get good results. With the continuous research of many experts and scholars in the field of artificial intelligence machine learning, Hinton, an internationally renowned expert on machine deep learning, proposed in 2006 that multi-layer hidden layer can learn more features, be closer to human cognition, and have a more essential cognition of data, which is conducive to the identification of things (Hinton et al., 2006). The proposed multi-layer convolutional neural network model based on artificial intelligence and machine learning has greatly promoted the rapid progress of related image processing applications. But there are still many problems and difficulties in the utilization of the image processing training through the network and the relevant algorithm models, such as the limit of the number of neural network layers (Hinton and Salakhutdinov, 2006), the problem of gradient disappearance (In deep neural networks, the gradient gradually decreases during backpropagation and eventually becomes very close to zero) in the network structure layer and excessive coupling in image processing training (He et al., 2014). To solve these problems, many scholars have optimized the relevant algorithms and structures, so as to improve the ability of image processing. Niu and Suen (2012) designed a hybrid CNN-SVM model, combined the automatic retrieval ability of CNN with the classification and recognition ability of SVM, and carried out an experiment on mnist digital database. The recognition rate of this model reached 99.4%. Krizhevsky et al. (2012) in 2010, by increasing and training convolution layers of convolutional neural networks and using regularization mode, effectively reduced over-fitting of fully connected layers. In 2012 imagenet lsvrc-2012 contest, The further optimization of convolutional neural network reduced the error rate of image processing to 15.3%, much higher than the second place of 26.2%. In the following period, image processing using artificial intelligence and machine learning has become a research hotspot of many technology companies. Google, Facebook and China's Baidu have joined in. Google proposed GoogleNet model (Szegedy et al., 2014) in 2014. In the aspect of image processing, the number of parameters is reduced to only 5 million while the processing algorithm is guaranteed to be the number of network layers. In 2019, MobileNet V3 (Howard et al., 2020) proposed by Google, not only guarantees the image processing accuracy, but also reduces the model parameters to only 3.1% of VGG-16, greatly reducing the amount of computation.

With the continuous development of artificial intelligence and machine learning, combined with the continuous progress of

current computer hardware facilities, the application of artificial intelligence and machine learning in image processing will become more and more intelligent, precise and efficient in the future. Compared to the other reviews, In this paper, the algorithm models commonly used in artificial intelligence and machine learning in image processing, the principles and advantages and disadvantages of different network models, and based on these models in various fields of application and future development are reviewed.

2 Common algorithm model

2.1 Artificial neural network

Artificial neural network (ANN) (Schapire, 1999; Dreiseitl and Ohno-Machado, 2002; Juan and Valdecantos, 2022; Dragović, 2022) is an algorithm model based on the simulation of brain neurons and neural network structure and function, which abstracts, simplifies and simulates the brain tissue structure and operation mechanism. It is an algorithm model that simulates the structure and function of biological nervous system. The model is based on multiple neurons, as shown in Figure 1, and consists of three parts: input layer, hidden layer and output layer (Egrioglu et al., 2022; Yusri et al., 2018). The hidden layer plays an important linking role in the model. There can be multiple hidden layers in a model and the number of neurons in the hidden layer is uncertain. When the neurons in the model receive the external input information, the information is stored and summed. The stored information is summed and the calculated value is output after the activation function is processed. The principle is shown in Figure 2 and Equation 1. When the output value is 1, the neuron becomes excited. Conversely, it is in a state of inhibition. At the same time, when the neuron sums the data, it uses different weights to sum the calculation. In Equation 1, $W_1, W_2, W_3, W_4 \dots$ are the weight coefficients of each input data. At the same time, according to the different types of information processed by the artificial neuron model, different activation functions are used (Yang, 2006).

The artificial neural network is trained on the basis of input parameters and output results. Through a lot of training, the neural network model completes the data processing according to the predetermined expectations of the designer. The training and learning methods for neural network models are divided into supervised learning and unsupervised learning (Wu and Feng, 2018; Kasabov et al., 2016). People can constantly adjust the learning algorithms in the neural network according to their own needs. Generally, the weights and deviations of the parameters in the neural network are adjusted.

Therefore, it can be seen that the artificial neural network model is essentially a complex artificial network structure with distributed storage, parallel processing, self-learning, and self-organization characteristics, which is formed by using mathematical algorithms to connect a large number of neuron processing units according to certain rules. This structure makes the neural network structure not only have the ability to high-speed search for optimization solutions, but also can complete innovative self-learning and environmental adaptation functions. The application of this advantage can effectively meet the needs of image processing, especially in the image compression processing link, using different

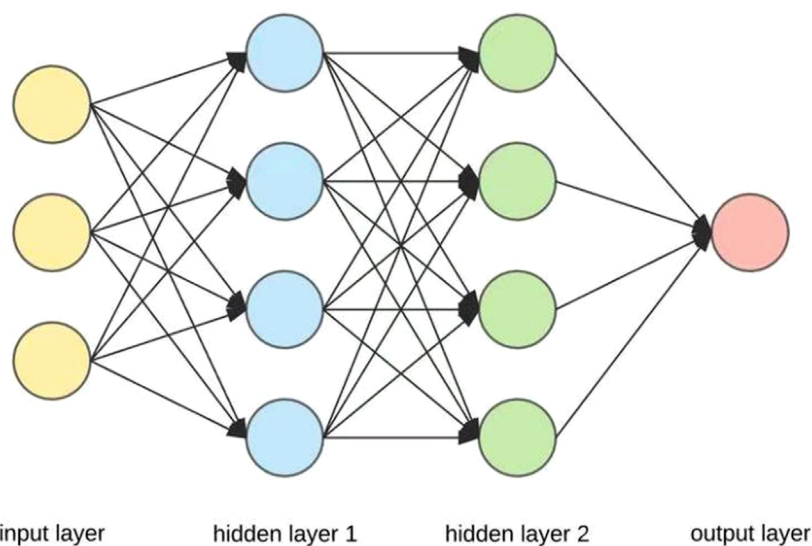


FIGURE 1 Artificial neural network model.

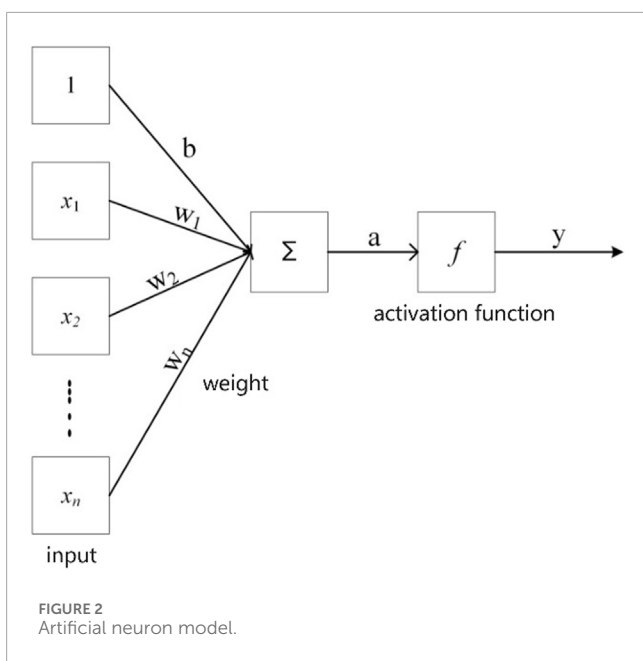


FIGURE 2 Artificial neuron model.

nodes and levels. Set different numbers of nodes to solve the characteristics of fewer data transmission level nodes, and ensure the rationality and effectiveness of data processing. At the same time, it can also greatly improve the data storage space and implement efficient transmission through the network to achieve the purpose of saving image space and improve the accuracy of image restoration to a certain extent.

Although artificial neural networks have many advantages, they inevitably have their own weaknesses: the most serious problem is the inability to explain their reasoning process and reasoning basis, and the neural network cannot work when the raw data is insufficient. The theory and learning algorithm of neural networks

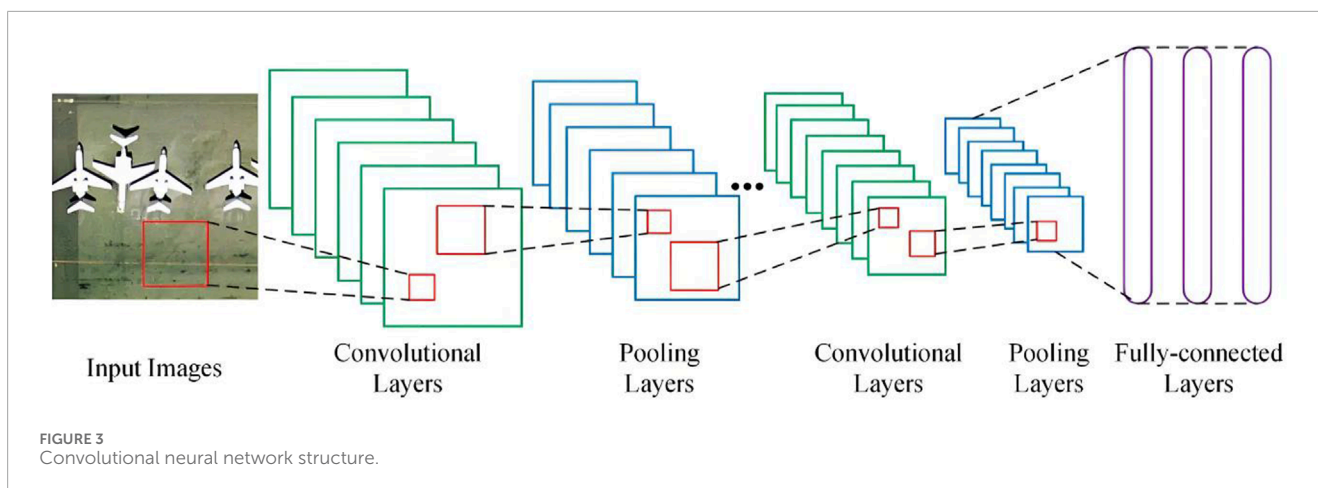
need to be further improved and improved.

$$y = f(a) = \begin{cases} 1, & \sum_{i=1}^n x_i w_i \geq \text{threshold value} \\ 0, & \sum_{i=1}^n x_i w_i \leq \text{threshold value} \end{cases} \quad (1)$$

2.2 Convolutional neural network

Convolutional neural network (Ulku and Akagunduz, 2019) is a further deep learning based on artificial neural network. In machine learning, convolutional neural network is a deep feedforward artificial neural network, which has been successfully applied to image recognition and large-scale image processing.

As shown in Figure 3, the composition of convolutional neural network mainly includes the following parts: input layer, activation function, convolutional layer, pooling layer, fully connected layer and output layer (Yuan et al., 2019; Zhou, 2018). In the input layer, image data is input, and each image is represented by pixel value matrix. Convolutional layer is the core component of convolutional neural network, and most of the work of image processing is completed in the convolutional layer. After the image data is passed from the input layer to the convolutional layer, the convolutional layer extracts the features of the input image data, and a group of convolution sub-composed of the convolutional layer learns different image features of the image data. Among them, since the input data is nonlinear, activation functions are introduced into the convolutional layer to transform the learning features of convolutional neural networks into nonlinearization. Commonly used activation functions include Sigmoid function Equation 2 and Tanh function Equation 3 (Adem, 2022; Zhang et al., 2023). The pooling layer reduces the dimension of the input data image by imitating human visual system, and represents the image with higher level features. At the same time, the image pixel matrix can be shrunk to efficiently shrink the input image data into a matrix, and the features of different regions in the image can be



aggregated for statistics, so as to reduce the excessive sensitivity of the convolution layer to the image position, reduce the parameters in the subsequent full connection layer, and improve the calculation speed (Zhou et al., 2022). In actual training, common Pooling methods include Average Pooling (Lai et al., 2017), Spatial Pyramid Pooling (He et al., 2015) and Max Pooling (Bera and Shrivastava, 2020). In addition, the scale invariance and rotation invariance of the convolutional neural network model can be improved through the pooling layer operation to avoid the phenomenon of overfitting. With the improvement of various pooling methods, the time of model training is greatly reduced. The fully connected layer is mainly used to integrate the features transferred by the conversion of the convolutional layer and the pooling layer, and the output layer can judge the category of related images to achieve the goal of image recognition. Convolutional neural networks include one-dimensional convolutional neural networks, two-dimensional convolutional neural networks and three-dimensional convolutional neural networks. One-dimensional convolutional neural networks are often used in sequential data processing. Two-dimensional convolutional neural networks are often used in image text recognition. 3D convolutional neural network is mainly used in medical image and video data recognition. Convolutional neural network technology has been widely used in image processing. According to the specific requirements of image processing, many convolutional neural network algorithm models are proposed for image processing applications. There are mainly the following:

$$f(x) = \frac{1}{1 + e^{-x}} \quad (2)$$

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (3)$$

2.2.1 AlexNet

AlexNet model (Krizhevsky et al., 2012) is shown in Figure 4. It is the first large convolutional neural network model with image processing proposed by Hinton research group in 2012 in ImageNet's large visual recognition challenge, and the error rate of this model algorithm is only 15.3%. AlexNet model uses multiple Gpus for data training, which greatly changes the disadvantages of traditional single GPU training algorithm. At the same time, ReLU

is used as activation function to improve the convergence speed of convolutional neural network. In order to solve a large number of network parameters to better regularize, thus avoiding the problem of overfitting, the AlexNet model uses the Dropout layer. Finally, the model further enhances the learning ability of CNN by means of parameter optimization.

2.2.2 2GoogLeNet model

GoogLeNet model (Szegedy et al., 2014) was proposed by the Google team in 2014. The number of layers of the GoogLeNet model can reach 22, which can greatly improve the capability of convolutional neural network. Compared with the VGGNet model, when the number of network layers is further increased, the number of related parameters in the GoogLeNet model is greatly reduced to less than 1/3 of the number of parameters in the VGGNet model and 1/15 of the original AlexNet model, so the requirements on computer hardware resources are greatly reduced. Although the performance of the model can be significantly improved simply and directly with the increase of the number of network layers, the network model becomes larger with the increase of the number of network layers, which is prone to gradient dispersion and overfitting. Therefore, the Google team proposed the sparsely connected Inception module to reduce the number of parameters, improve the utilization rate of computing resources, and overcome the problem of repetitive and complex information. On the basis of this module, Inception V2, Inception V3 and Inception V4 models are successively introduced, as shown in Figure 5. The GoogLeNet model is further improved.

2.2.3 VGGNet model

VGGNet ((Visual Geometry Group) (Gao and Mosalam, 2018) convolutional neural network model on the basis of AlexNet model and LeNet model uses a convolutional layer stacked with 3×3 small convolution kernels to further improve the depth of the network and enhance the performance of the model. Compared with AlexNet model, the computational cost of VGGNet model is further reduced due to the reduction of convolution kernel. VGGNet model can set more network layers, can be set to 19 layers, through the increase of the number of network layers, the network performance can be significantly improved. However, as there are many convolutional

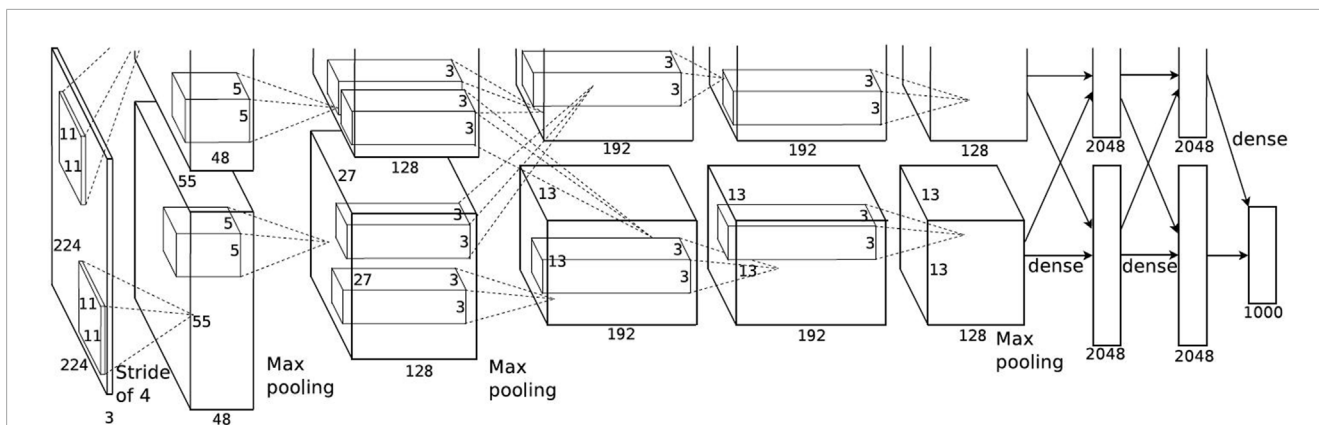


FIGURE 4 AlexNet model (Krizhevsky et al., 2012).

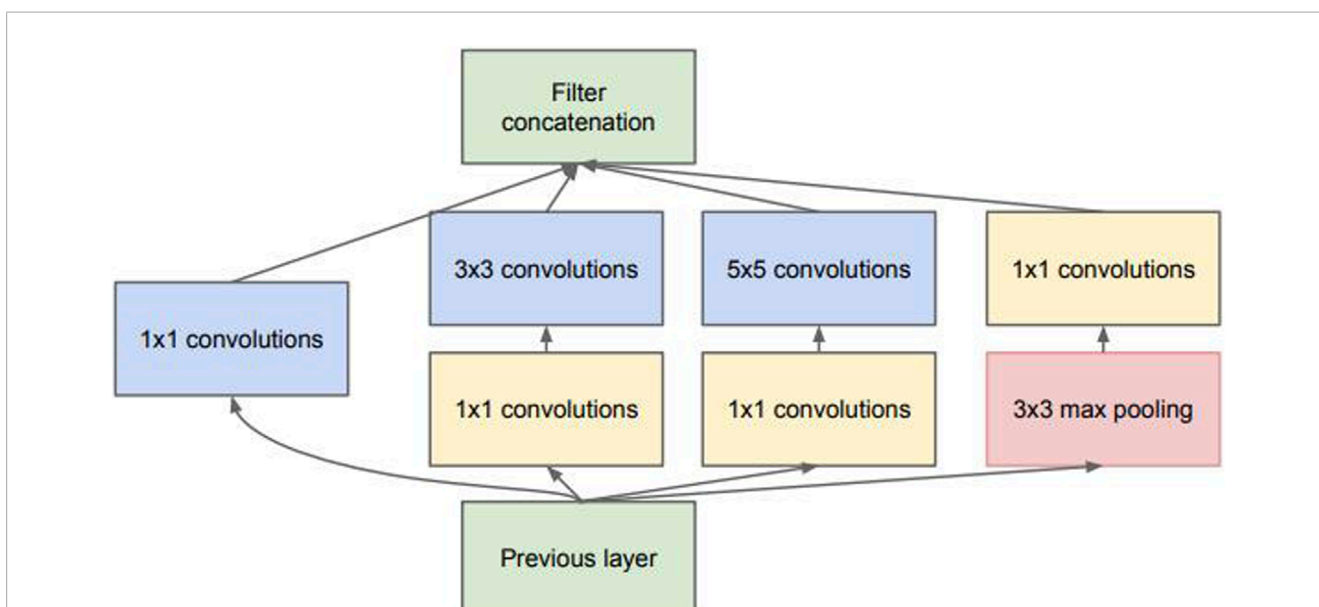


FIGURE 5 Conceptual model of Inception module (Szegedy et al., 2017).

layers stacked in the VGGNet model, many parameters of the model are used, which has higher requirements for the system in use. Commonly used VGGNet models include VGGNET-16 structure and VGGNET-19 structure.

2.2.4 ResNet model

ResNet model (residual networks) is the residual convolutional neural network model proposed by He et al. (2016) in 2015, which is based on the convolutional neural network. The ResNet model is mainly designed to solve the problem that although the number of layers of the network in the VGGNet model and GoogleNet model is increasing, the performance of image feature extraction and processing is getting better and better, the optimization of the network is becoming more and more difficult with the increase of the depth of the network. In addition, simply increasing the number of layers of the network has no obvious effect on model training,

and the error rate of the model will also increase. ResNet model uses residuals block to continuously approximate residuals mapping $F(x)$ to original mapping $H(x)$, where $H(x) = F(x) + x$, where x is input data. The specific structure is shown in Figure 6. The ResNet model further deepens the depth of the network to 152 layers through the stacking of residuals, which greatly enhances the image processing ability. However, the above problems of difficulty in optimization and gradient disappearance still exist.

2.2.5 U-Net model

U-net model is a neural network model based on the expansion and modification of the full convolutional network proposed by Olaf Ronneberger in 2015. Because the model architecture presents the shape of the letter U, it is called U-Net (Ronneberger et al., 2015). As shown in Figure 7, the network consists of two parts: a shrinking path to obtain context information and a symmetric expanding path to

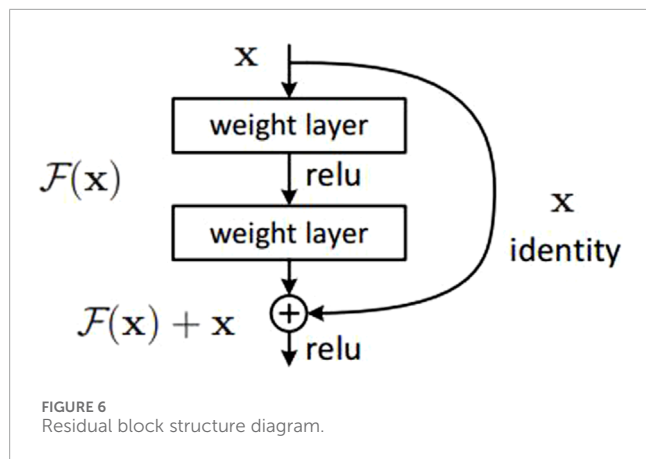


FIGURE 6 Residual block structure diagram.

accurately locate the information of the identified object. Firstly, the up-sampling convolution process of $2 * 2$ is carried out for the identified object each time through the path on the left, and the number of corresponding feature channels is reduced to half of the previous one. Then, the down-sampling is carried out through the right network and the features in the up-sampling process are connected. The whole network carried out 19 convolution operations, 4 pooling operations, 4 up-sampling operations, 4 clipping and copying operations. The convolution layer uses the mode of “valid, padding = 0, stride = 1” (It refers to the model where the input data is not filled in during the convolution process, and the size of the output feature graph shrinks according to the size and stride size of the convolution kernel) for convolution, so it can be seen that the final output image is smaller than the original image. To get an output image of the same size as the original image, you can use the “same” mode during the convolution operation. This algorithm has a good performance in small target image recognition, especially in the field of medical image. At the same time, compared with traditional deep learning, which requires more data training sets, U-Net can obtain better segmentation effect after training with a small number of data sets. This is because the model classifies and processes each pixel of the target image to obtain higher segmentation accuracy. In addition, the image segmentation speed of U-Net model after training is also very fast. However, due to the use of effective convolution in U-Net model, the difficulty and universality of the model design are increased. At the same time, in some small target image recognition and segmentation tasks, the problems of small target loss and inaccurate edge of fuzzy object sometimes appear.

Due to the efficient performance of U-Net model, many scholars have successively proposed UNet++ (Zhou et al., 2018), nnU-Net (Isensee et al., 2021), UNet 3+ (Huang et al., 2020) and other classical neural network models on this basis for different details and needs, which have been continuously improved and developed.

2.2.6 CliqueNet model

CliqueNet (cluster network) was proposed by Yang et al. (2018) in 2018. CliqueNet model adopts cluster connection and network iteration mechanism to connect any two layers in the same group forward or reverse. Constantly updated. Therefore, the CliqueNet model, as shown in Figure 8, can maximize the transmission efficiency between feature layers and effectively improve the

accuracy of the network through feature re-calibration of channels and spatial dimensions in the network conversion layer.

2.3 Support vector machine

The support vector machine (SVM) model (Chauhan et al., 2019; Rahat et al., 2019) was proposed by Vapnik et al. in the 1990s. The SVM model is based on the relevant theories of statistics and the principle of minimum structural risk. Based on the limited sample information, Seek the best compromise between the complexity of the model (i.e., the learning accuracy of a specific training sample) and the learning ability (i.e., the ability to identify any sample without error) to obtain the best ability. Through the support vector machine model, the relevant features of the image are taken as the input parameters of the classification of the support vector machine through certain algorithms, and as the classifier of the core component of the support vector machine, the image data is trained, and the given data is divided into two categories according to the rules, linear classifiable and nonlinear, and the optimal classification surface is found in all the classification surfaces. A classifier model with minimal structural risk is constructed to realize the target of image recognition.

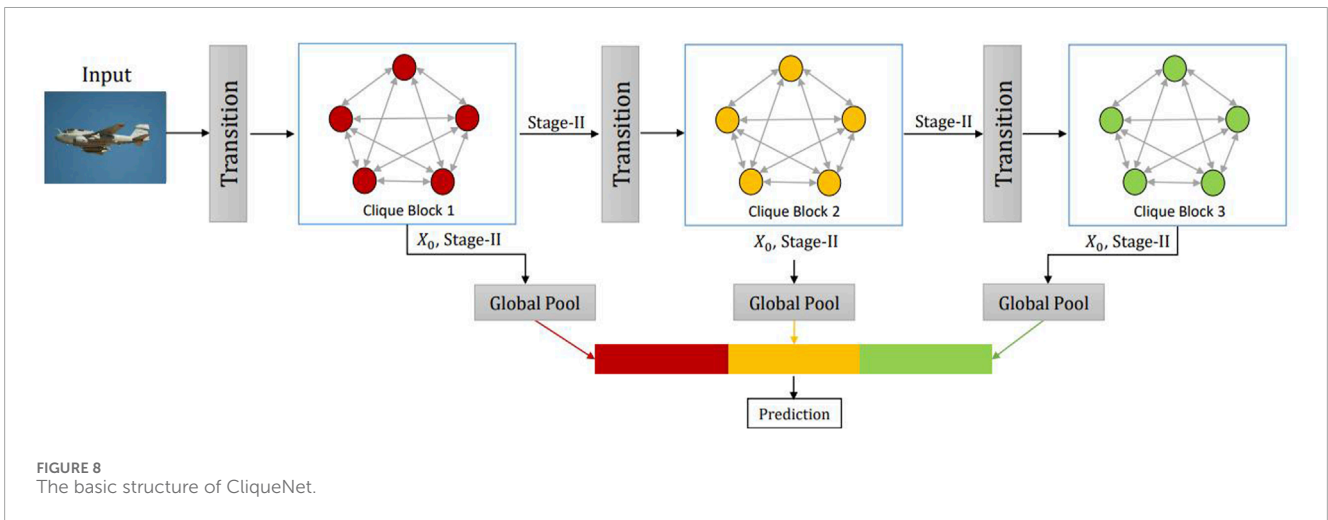
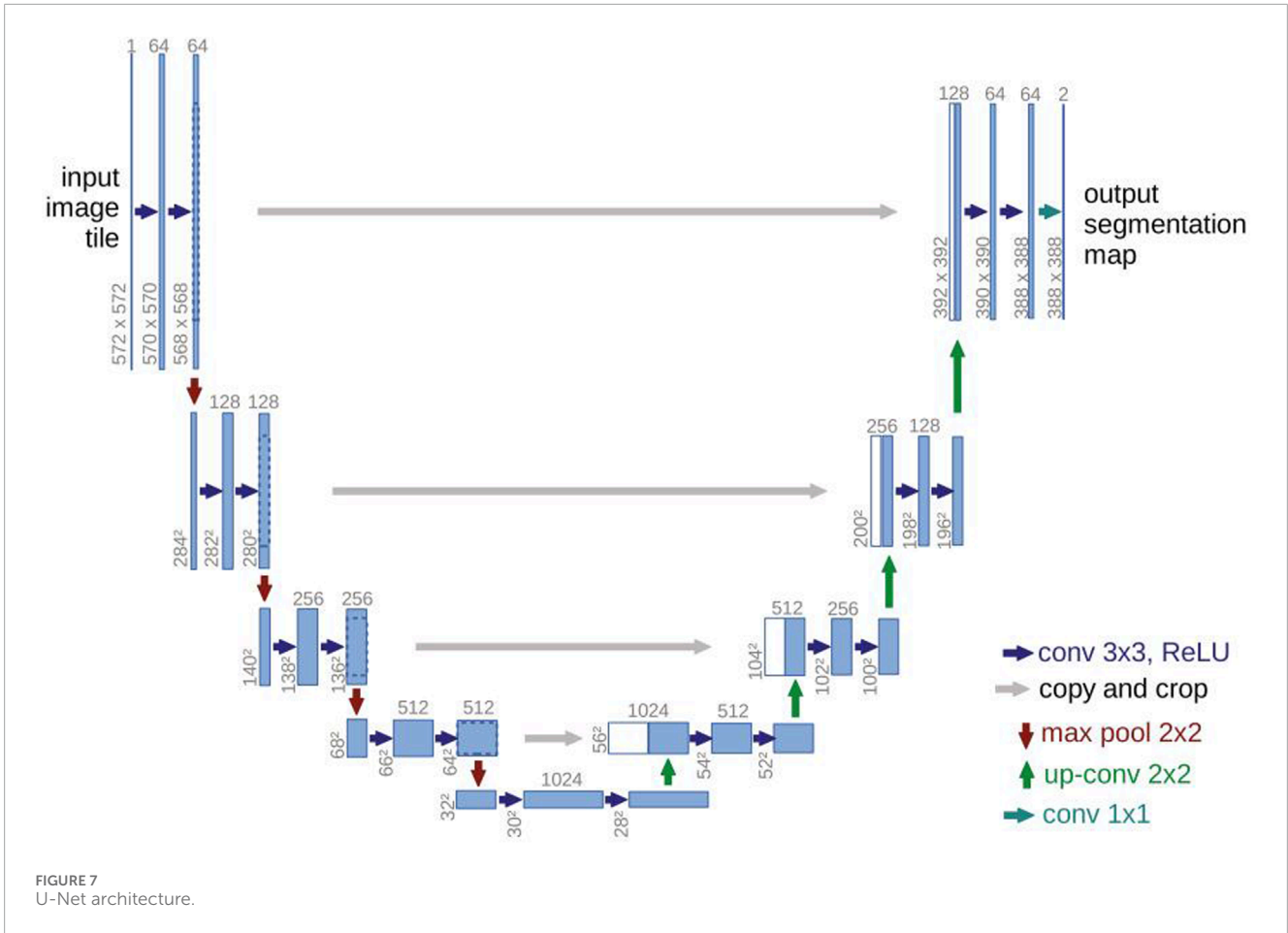
2.3.1 Linear support vector machine (SVM)

For linear classifiable problems, the optimal classification surface is directly found with a given sample. Function $G(X) = ax + b$ is used to construct the classification surface, where a represents the direction vector of the plane vector, and b represents the offset of the plane. The value of the function $G(X)$ is shown in Equation 3. Using the selected sample number m as the sample space, the training set of the sample is divided into $(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)$, the sample is represented by (x_i, y_i) , x_i represents multidimensional linear fractional data, y_i represents the category where x_i is located. After solving the corresponding linear classifier by using the equation, the maximum classification interval can be solved, and the optimal classification surface can be obtained, so as to process the corresponding data.

$$G(X) \begin{cases} = 0; \text{Points on the classification plane} \\ < 0; \text{Points with sample type of } -1 \\ > 0; \text{Points with sample type of } 1 \end{cases} \quad (4)$$

2.3.2 Nonlinear support vector machine

In daily life, linear unclassifiable data are mostly used. It is impossible to classify data accurately through linear classifiers. When the problem is linearly non-separable, there will be isolated points in the given data sample, and the linear classifier will classify the isolated points incorrectly. In order to solve this problem, we add relaxation variable to the solving analysis of classification function, in which relaxation variable refers to the distance between the isolated point and the plane where the isolated point is located. In solving the classification function, a constant is specified as a penalty factor to limit the value of the relaxation variable. The larger the value of the relaxation variable is, the farther the distance between the isolated points and the correct classification plane will be. The larger the value of the penalty factor is, the more points can be correctly classified. In order to solve the nonlinear classification problem, we need to introduce kernel function, through which the low-dimensional



sample space is transformed into a linearly separable high sample space by mapping relation. The transformation relation we use is kernel function. Under the transformation of kernel function, the two-dimensional indivisible sample data is converted into linear fractional data in three-dimensional space, so that we can find the most classification surface, and then find the classification function. In support vector machine model, kernel function as the core determines whether the classifier can carry out better reasonable classification. The

types of kernel functions commonly used mainly include polynomial function, Gaussian radial function and S function.

2.4 Swarm intelligence algorithm model

Swarm intelligence (Beni and Wang, 1993) was first formally proposed by Beni and Wang. This algorithm is mainly proposed

based on the principle of intelligent cooperation of biological groups such as ant colonies and migratory bird colonies in nature. Swarm intelligence algorithm through the principle of a large number of individuals composed of group behavior research, the group behavior modeling, set up the corresponding rules, and then form the corresponding algorithm. According to different starting points of research objects (Bhardwaj et al., 2022), swarm intelligence algorithms can be divided into collective intelligence represented by ant colony algorithm, which regards the research group as a group of agents, and particle swarm optimization algorithm, which regards the members of the research group as individual particles. In the aspect of image processing, swarm intelligence algorithm is mainly combined with image segmentation. Many scholars continue to optimize on the basis of ant colony algorithm and particle swarm optimization algorithm to ensure the accuracy of image segmentation. Under the premise of image segmentation, image segmentation processing speed is further improved.

2.4.1 Ant colony algorithm

The principle of Ant Colony Optimization (ACO) (Zhang and Dahu, 2019) comes from the fact that when ants go out to look for food, they can always find the shortest route back to the nest by avoiding the corresponding obstacles. Based on the analysis of this activity of ants looking for food, scholars proposed Dorigo et al. proposed the ant colony algorithm (Dorigo and Di Caro, 1999). The algorithm was first applied to the traveling salesman problem, then ant colony algorithm has been used in many types of research, and has been continuously optimized. The main principles of the algorithm (Gutjahr, 2000; Sen Sarma and Bandyopadhyay, 1989; Ilie and Bădică, 2013) are shown as follows:

Firstly, the positions of n ants are randomly assigned to complete the initialization. At first, on the way ants set out to find food, the pheromone concentration on each path was the same, with the first ants, the second... to the nth. The probability of each ant choosing the path of departure is shown in Equation 4.

$$p_{ij}^k(t) = \begin{cases} \frac{[\tau_{ij}(t)]^\alpha [\varphi_{ij}]^\beta}{\sum_{s \in allowed_k} [\tau_{is}(t)]^\alpha [\varphi_{is}(t)]^\beta}, & j \in allowed_k \\ 0, & Otherwise \end{cases} \quad (5)$$

In the formula: $p_{ij}^k(t)$ represents the probability function of the path randomly selected by the ant; τ_{ij} represents the pheromone representing the edge of the path; φ_{ij} represents the heuristic factor of the ant from the starting point to the end point, $\varphi_{ij} = 1/d_{ij}$; d_{ij} represents the distance from the starting point to the end point; $allowed_k$ is the set of locations that the ant can access

Then, after a period of time, all the ants that had gone out looking for food returned to their original locations. At this time, the pheromone concentration of each ant going out to find the road will change. After the ant colony's exploration, the nearest path will be retained, and the pheromone concentration at the edge of the path will be correspondingly high (Smith et al., 2014). Of course, the pheromone concentration left on the edge of the road where the ants go out to search for food will also evaporate over time. The changes of pheromone concentration on the edge of the road are shown in Equations 5, 6.

$$\tau_{ij} = (1 - \rho)\tau_{ij} \quad (6)$$

$$\tau_{ij} = \tau_{ij} + \sum_{k=1}^m \Delta\tau_{ij}^k$$

In the equation: ρ is the coefficient of pheromone volatilization; $\Delta\tau_{ij}^k$ is the pheromone released by the k th ant at the edge of the path from the starting point (i) to the end point (j);

Finally, the ants returning to the original location prepare for the next search for food, constantly optimize the path, and the pheromone concentration on the edge of the path is constantly changed, which should also take into account the volatilization of pheromones on the edge of the path. Finally, find the best path in the constant loop.

In image processing, an ant can be regarded as a pixel in the image. The process of ants going out to search for food is just like image segmentation in image processing. By using the positive feedback mechanism, ants can find the best path principle after searching for food, and the image segmentation is more accurate. With the further study of ant colony algorithm, Qin et al. (2019) improved the initial clustering center by updating the parameter of pheromone concentration and combining with the threshold segmentation method of image, thus further improving the accuracy of image processing and segmentation. El-Khatib et al. (2019) proposed a new segmentation method for image processing by combining the advantages of k-means algorithm with ant colony algorithm, which is obviously superior to the effect of single algorithm.

2.4.2 Particle swarm optimization algorithm

Particle Swarm Optimization Algorithm (PSO) is a swarm intelligence algorithm proposed by Kennedy and Eberhart through the swarm behavior study of bird foraging behavior (Kennedy and Eberhart, 1995). The principle of the algorithm is derived from the fact that cooperative hunting birds communicate with each other based on the experience of the group members to optimize the search pattern and improve the efficiency of searching for food. We regard each bird in the flock as a particle, and the whole flock is equivalent to a particle swarm. Particle swarm is in a space of random solutions, and particles share their optimal solutions with each other. After optimization function processing, there is an optimal solution for a single particle and a global optimal solution for the whole particle swarm.

The specific principle of particle swarm optimization algorithm (Marini and Walczak, 2015; Tang et al., 2021) is shown in Equations 7, 8:

$$v_{t+1} = c_1 v_t + c_2 r_1 (p_{i,t} - x_t) + c_3 r_2 (p_{g,t} - x_t) \quad (7)$$

$$x_{t+1} = x_t + v_{t+1} \quad (8)$$

In the formula: v_t and x_t represent the velocity and position of the particle at time t respectively; $p_{i,t}$ represents the optimal solution of a single particle before time t. $p_{g,t}$ represents the global optimal solution of the particle swarm before time t; c_1, c_2, c_3 represent the cognitive coefficient of particles; r_1, r_2 are random parameters in the range of [0, 1].

In Equation 6, $c_1 v_t$ is the velocity vector of the particle at this point in time, and $c_2 r_1 (p_{i,t} - x_t)$ represents the vector from the current value to the optimal value of the particle, which is affected by the cognition coefficient. $c_3 r_2 (p_{g,t} - x_t)$ represents the vector from

the current value to the global optimal solution, which represents the cooperation and information exchange ability of particle swarm.

In the particle swarm optimization algorithm, each particle has only two attribute parameters speed and position, which also makes the parameter setting of the algorithm simple and the processing speed fast. Since the particle swarm optimization algorithm was proposed, scholars from various aspects have made continuous improvements, mainly including the improvement of the algorithm itself (such as parameter change, algorithm structure optimization) and the fusion of particle swarm optimization algorithm with other algorithms. For example, Kangling (2008) used fuzzy entropy as the target and particle swarm optimization algorithm to solve the problem of selecting the optimal threshold in image processing. Zhang and Lim (2020) combined deep neural network with particle swarm optimization algorithm to make the search process of retinal image segmentation more diversified.

3 Applications in different fields

3.1 Application of medical image processing

Artificial intelligence and machine learning, as auxiliary means in medical imaging, can greatly improve the working efficiency of doctors (Kumar et al., 2022). At present, the application model based on convolutional neural network (CNN) algorithm has been used in diagnosis of various images in the medical field. Common MRI image analysis, CT image analysis, ultrasonic image analysis and X-ray image analysis.

Aiming at the problems of medical image quality and imaging speed, Long et al. (2015) used the complete convolutional network to process images and replaced the fully connected layer in the convolutional network by using the restored convolution and maximum pool in the complete convolutional network to make the output image become full resolution. Improve the imaging quality and speed of medical imaging. In specific medical treatment, most medical images are mainly used to image the detection parts of the human body through radioactive substances. The main principle is to make use of the differences in density and thickness of various tissues and organs of the human body and the differences in the amount of radiation absorption, so as to change the distribution of transmitted rays and finally present them through a series of transformations. The larger the amount of radioactive material is, the better the imaging effect will be, but it will cause certain damage to human experience. Therefore, Lyu et al. (2021) proposed a machine learning model based on convolutional neural network, which can transform medical imaging under low dose radiation into medical imaging under normal dose, so as to ensure imaging quality and reduce radiation damage to patients. In daily diagnosis, doctors often need to compare and analyze medical images. At this time, different images need to be mapped to an image. Traditional methods often need to consider measurement methods, similarity, image-based registration and other parameters, which is time-consuming and labor-intensive. However, the creation of a learning model based on artificial intelligence and machine learning can effectively solve the image registration problem. For example, Cheng et al. (2018) proposed an artificial intelligence and

machine learning model for multi-modal CT-MR registration of skull images based on DNN architecture. In addition, in medical ultrasound image imaging, the scattered echo of ultrasound will lead to the imaging and human organs and tissues are not clearly distinguished, affecting the judgment of doctors. Trinh et al. (2011) proposed an adaptive mean filtering ultrasonic image noise reduction model using support vector machine (SVM). This model uses the characteristics of support vector machine classification to distinguish noise from non-noise in ultrasonic images, and finally combines with mean filtering to denoise the images, so as to better preserve the edge features of human tissue region in imaging.

The use of artificial intelligence and machine learning in image processing related diseases has been a great success to this day. The artificial intelligence system developed by Harvard Medical School has been able to identify cancer cells in pictures of breast cancer cases with an accuracy of 92%, and combined with artificial pathological analysis, its diagnostic accuracy can reach 99.5% (Luchini et al., 2022). With the further application of artificial intelligence and machine learning to medical image processing, image analysis and disease diagnosis will be more efficient and intelligent.

The industry has recognised the potential applications of employing artificial intelligence to create annotated data sets of medical photographs, as the need for medical AI scenarios to be launched grows. In certain studies, real-world clinical data can be assigned using natural language processing and image cognitive models; this is accomplished through the use of the people-in-the-loop (human-in-the-loop) method, iterative building based on clinical data annotation, standardisation, and large-scale data; in the future, it is anticipated that the data from medical image analysis and medical research methods will be combined.

3.2 Application of facial image recognition processing

In general, there are three main components to a facial expression recognition system: feature extraction and discriminant analysis, face detection and normalization, and classification and validation. Because we can only significantly increase the performance and impact of the classification algorithm by precisely extracting the representative feature set, face feature extraction has the most influence on face recognition and face expression recognition in the three steps mentioned above. Numerous techniques for extracting face features have been developed; the wavelet transform is the most well-known and frequently utilized one (Wu et al., 2010). The wavelet transform's capacity to locate in the time and frequency domain made it a valuable tool for feature extraction in the past. It is limited to capturing the point singularities in the image; it is unable to capture the facial feature that sets the face apart, which are the curves and lines. Curve transformation-based pattern recognition has been the subject of numerous investigations. Tang and Chen (2013) proposed an improved facial image recognition method by using particle swarm optimization (PSO) and improved support vector machine model (SVM), which further improved the accuracy and robustness of facial image recognition. In 2014, Sun et al. (2014a), Sun et al. (2014b) proposed DeepID based on convolutional neural network. The accuracy of this

algorithm in facial image recognition by CeleFace reached 97.45%. After improving the depth of the algorithm and increasing the number of layers, DeepID2 and DeepID2+ were introduced, and the recognition rate of facial images reached 99.47%, surpassing the level of human eyes. With the wide application of facial image recognition, it becomes more and more convenient to use facial image information to identify and search individuals. Because of the convenience of facial information collection, simple devices such as cameras or video cameras can be used to capture facial information. [Ragul and Cloudin \(2021\)](#) used a large number of facial image databases to train convolutional neural networks, and proposed a face recognition method based on density network for the recognition and confirmation of criminals. Compared with the traditional single fingerprint confirmation, this method is faster and more efficient. In recent years, in some fields with high security requirements, the application of artificial intelligence to analyze and process facial images has become more and more popular in the field of security. [Singh et al. \(2021\)](#) proposed an imaging system technology based on rapid development. Through the processing and analysis of acquired facial images, real-time monitoring of continuously obtained facial images can be realized and the changing trend of acquired facial image data can be mastered under the condition of ensuring physical and digital security. In view of this feature, the paper “Real-time Detection of Suspicious Events of Bank Automated Teller Machines Based on RGB + D and Deep Learning” argues that with the increase of illegal activities of automatic teller machines, real-time monitoring can be utilized to achieve real-time analysis of facial images, and some abnormal behaviors can be recorded in time to make responses ([Khaire and Kumar, 2021](#)). To get rid of the problem of relying solely on human supervision by security guards.

Not only is AI figuring out how to collect more available data, but it also needs to figure out how to create more accurate models. In the mode of continuous improvement, there are still many problems waiting for us to solve, especially the deep learning face recognition will be an improved mode. For example, in the case of convolutional neural network, how to separate a higher requested image feature, how to optimize the confirmation rate and elasticity of pictures, and in the case of sublayers of face recognition, the choice of parameters still needs further research.

3.3 Application of agricultural image processing

With the development of artificial intelligence and machine learning, especially the great success of deep machine learning based on convolutional neural network in image processing, the research on agricultural image processing using artificial intelligence and machine learning models is getting deeper and deeper. At present, agricultural image processing using artificial intelligence and machine learning is mainly focused on crop growth monitoring, agricultural product quality detection and crop diseases ([Khan et al., 2022](#)).

In terms of crop growth detection, artificial intelligence and machine learning can be used to extract features such as color and shape of leaves, rhizomes and fruits in the process of crop growth. Through the design and calculation of relevant algorithms, the growth health status and maturity of crops can be judged,

greatly improving the efficiency of agricultural production and management. However, the performance of the plant disease classification system mainly depends on the feature extraction efficiency and the type of classifier used for training. These features correspond mainly to the leaf disease symptoms. However, these techniques provide good performance in small datasets, but their performance decreases when used in large datasets. Machine learning techniques such as random forest (RF) and artificial neural networks (ANN) can provide improved results, but with long parameters and computation time. The processing of these architectures is also very time-consuming and complex. In this regard, [Sahu and Pandey \(2023\)](#) proposed a new hybrid random forest multi-class SVM (HRF-MCSVM) model for the detection of plant leaf diseases, which effectively solved the problem of time-consuming and complex computation in the large data volume of traditional random forest (RF) and artificial neural network (ANN). At the same time, 54,303 healthy and diseased leaf images were selected for testing, and the recognition accuracy of HRF-MCSVM was 98.9%. In terms of crop diseases and pests, it is difficult to determine the accuracy of disease classifiers and small marker data sets (10–800 images). Many scholars use convolutional neural networks to train corresponding machine learning models with a large number of crop diseases and insect pests images as databases. [Brahimi et al. \(2017\)](#) trained the proposed convolutional neural network model based on 14,828 images of 9 kinds of tomato leaf diseases, and the accuracy of the model for disease recognition of tomato leaves reached 99.18% at last. [Liang et al. \(2019\)](#) used convolutional neural networks to build a deep learning model for feature extraction and classification of rice blast. Compared with traditional recognition methods, this model had stronger recognition ability and the accuracy of classification of rice blast was up to 95%. It has greatly improved people's ability to prevent crop diseases. Food safety issues are being mentioned more and more. In terms of quality monitoring of agricultural products, relevant scholars use the structural composition of agricultural products and food to occupy a specific amount of light, that is, different structures will show different characteristics of light absorption and scattering. This is because materials in different tissues absorb light at different rates. [Lorente et al. \(2013\)](#) used five emission light sources to detect citrus rot. Through the analysis of the image of citrus after light irradiation, Gaussian Ehrentz function was used to describe the backscattering characteristics. Using linear discriminant analysis as classification model, the detection accuracy reached 80.4% at 532 nm, and the detection accuracy reached 96.1% with the combination of five light wavelengths.

In recent years, the use of hybrid systems (such as neuroblur or image processing) combined with artificial neural networks is emerging. It moves toward more automated, more accurate systems that can take action in real-time. Further research is being conducted using more advanced tools so that traditional agriculture can move towards precision agriculture at a low cost.

3.4 Application of asphalt mixture CT image processing

In the present asphalt pavement design method, generally used is mainly Marshall design method and Superpave design system. In

practical engineering, these two methods are mainly based on the macro performance of asphalt pavement to study the composition of asphalt mixture pavement material design, so as to establish the relationship between the macro performance of asphalt pavement and asphalt mixture design. However, with the further study of asphalt mixture performance, it is found that the meso-structure of asphalt mixture plays a crucial role in the performance of asphalt mixture, and the meso-structure of asphalt mixture often affects the macroscopic performance of pavement. Therefore, many researchers transfer the study of asphalt mixture to the microscopic, with the help of CT technology and the development and digital image processing technology to study the internal structure of asphalt mixture for a lot of research.

After the scanned image is obtained from the asphalt mixture, the physical image is generally converted into a digital image and then the useful information is extracted from the digital image through computer processing. At the same time, the acquired image should be restored and enhanced, and then the acquired image should be processed by smoothing, sharpening, threshold segmentation and other methods, so as to improve the image quality. Finally, the image information is further studied to obtain useful feature parameters (Lorente et al., 2013). Especially after the image enhancement, the image segmentation method based on threshold is a common image segmentation method, the main idea is that different targets have different characteristics, such as color, gray level, contour, using the subtle differences between the features, select the threshold to divide into the target and the background. Threshold segmentation can realize image segmentation quickly. It is often used when the target and background are quite different, and it is difficult to recognize complex environment. The image segmentation based on this method will still have the error of aggregate boundary segmentation, and the adhesion of aggregate will occur, which will lead to the error of the identification of the number of subsequent sequels, and then affect the relevant calculation of subsequent parameters such as volume fraction.

With the progress of artificial intelligence and machine deep learning, artificial intelligence and machine learning are used in image segmentation. After inputting the image features to be recognized in the system, a stable and mature system is formed after a large number of image segmentation exercises by the machine, which can effectively segment the characteristics of aggregates in the image. For example, Dollar et al. (2006) proposed a BEL edge detection algorithm based on the PBT (Boosting Tree) classifier. Jin et al. (2022) introduced Inception's multi-block mechanism and DenseNet's dense connection mechanism to solve the problem of color difference in ore image recognition. Based on semantic segmentation network U-Net, and built a multi-module and densely connected U-Net deep convolutional network mining and rock image segmentation algorithm. Compared with traditional segmentation methods such as maximum variance method, cluster analysis, edge extraction, etc., this mining and rock image segmentation algorithm based on U-Net deep convolutional network has high segmentation accuracy and good segmentation effect, and can directly obtain the number of ore and rock blocks in the effect map, greatly reducing the amount of image follow-up processing and improving the processing speed. Future trends are likely to continue to focus on how to further improve the performance of U-Net and its variants, including but not limited

to developing more efficient training algorithms, introducing more jump connections to capture more contextual information, and exploring the application possibilities of U-Net in other areas.

3.5 Applications in the field of transportation

The application of artificial intelligence and machine learning has been well applied in many fields. Based on the analysis and modeling based on traffic big data, artificial intelligence and machine learning have also been applied in the field of transportation in various directions, bringing traditional transportation into the era of intelligent transportation. At present, the application of artificial intelligence and machine learning in the field of traffic mainly focuses on traffic flow prediction, traffic law enforcement, traffic accident prediction and road detection.

In terms of traffic management, traffic flow prediction is a part of an intelligent transportation system. Passengers and government departments can receive real-time traffic information and traffic decision basis to solve the congestion problem prediction through accurate and reliable traffic flow results. The sensor system installed on various road types provides traffic network traffic, speed, lane occupancy rate and other information, to realize short-term traffic flow prediction. How to improve the accuracy of traffic flow prediction has always been the focus of intelligent transportation system. Artificial intelligence algorithm and machine deep learning are used to put the corresponding model into the collected massive traffic flow data and images for training and learning. We constantly make the corresponding adjustments in the model training, and finally get the model that can predict the traffic flow. Medina-Salgado et al. (2022) proposed an improved KNN model considering the existing temporal and spatial correlations in the traffic network by adopting a new dynamic distance metric that uses physical and data characteristics to replace physical distance and is represented in the state matrix to simulate traffic conditions. Comparative testing with other intelligent algorithms on traffic data set of Liuliqiao District in Beijing shows that the proposed model shows higher accuracy, and the MAPE value is 2.96% higher than the original KNN model. Lin et al. (2020) proposed a traffic flow prediction model based on correlation vector machine by optimizing the parameters of combinatorial kernel function through genetic algorithm and particle swarm optimization, and verified the proposed model with real data of Whitemud Drive in Canada. Experimental results show that the proposed model has higher accuracy than other prediction models, and the optimized model significantly reduces the algorithm time. Ganji et al. (2022) developed an AADT prediction model based on vehicle detection and road feature extraction from Google aerial images. R-CNN detected vehicles in aerial images and captured vehicles and road features of more than 17,800 roads from Google aerial images. The method of extracting road area, road width and road direction is developed, and the time stamp of aerial image is captured. The results show that the proposed method has high accuracy. Although the traffic flow prediction model has made great development in recent years, behind the improvement of the prediction accuracy is a larger, more complex and more difficult to train deep learning model. The training and deployment of these models requires a lot of training

time and computing resources; more complex models mean more training data, limiting the generalization and deployment of models. Therefore, in order to apply to more application scenarios and fast delivery requirements, the design of light and high-performance traffic prediction model is one of the key breakthrough directions in this research field.

In terms of traffic law enforcement, the current traditional license plate positioning method cannot detect license plates in complex road conditions such as bad weather conditions and view point changes. Moreover, machine learning-based license plate positioning methods have difficulty in accurately locating license plate areas, and license plate positioning methods may mistakenly detect objects similar to license plates, such as billboards and road signs. As a result, some scholars have used artificial intelligence and machine learning to process images and videos based on the large number of photos generated by surveillance photography and photography. After the established model is put into the image database for training, it can effectively identify related illegal traffic behaviors and greatly improve the efficiency of traffic law enforcement. [Chen et al. \(2016\)](#) developed an adaptive sparse reconstruction method of vehicle behavior learning based on video surveillance system to distinguish normal and abnormal vehicle motion modes. In the experiment, CROSS, i-LIDA, Stop Sign and I5 data sets were used to verify the performance and effectiveness of the proposed method. The results show that the classification and anomaly detection accuracy of this method is better than those of naive Bayes classifier, k nearest neighbor, support vector machine and traditional trajectory learning method based on sparse reconstruction, which can provide some references for traffic management, public service and law enforcement. Due to the uniformity of license plates, this feature can be used to carry out effective image processing, so that the traffic management department can better discover the problems in vehicle operation, reduce the traffic accident rate, and improve the efficiency of traffic law enforcement. [Min et al. \(2019\)](#) proposed a new vehicle license plate location method based on novel YOLO-L and license plate pre-recognition. By using the K-means ++ clustering algorithm to select the optimal number and size of candidate frames and modify the structure and depth of the YOLOv2 model, license plates and similar objects can be effectively distinguished. Experimental results show that the proposed method achieves 98.86% accuracy and 98.86% recall rate, which is superior to the existing method. However, the current detection speed can be used in real time with the assistance of the GPU. Detection efficiency can be further improved by further reducing the detection area. We can study this further in the future.

Traffic accidents have become a very serious concern worldwide because it can lead to major injuries, death and property damage. Therefore, predicting the severity of traffic accidents is of great significance for the government traffic departments to formulate traffic safety policies, as it can reduce casualties and property losses. Furthermore, identifying the key factors affecting the severity of traffic accidents is crucial for choosing safety countermeasures and strategies to mitigate the severity of traffic accidents. At present, scholars use the existing traffic accident data statistics to model the road traffic accident based on the corresponding machine deep learning, and apply the model to the traffic accident prediction. [Alqatawna et al. \(2021\)](#) effectively predicted

traffic accidents by analyzing the data of traffic accidents in Spain from 2014 to 2017 and training the artificial neural network model, and the predicted results were close to the actual results of highway traffic accidents. [Yang et al. \(2022\)](#) proposed a multi-task model based on deep neural network framework to predict the severity of traffic accidents on the basis of deep neural network. The model also combines the factors of property loss that people ignore in traffic accidents. One of the future challenges of road traffic forecasting is to expand the scope of proposed models and predictions by integrating heterogeneous data sources, including geospatial data, traffic information, traffic statistics, video, sound, and emotion from social media, which many authors believe can improve the accuracy and accuracy of analysis and prediction.

4 Conclusion

With the continuous development and progress of artificial intelligence and machine learning, the application of artificial intelligence and machine learning in image processing is more and more extensive. Compared with the traditional commonly used finite element, discrete element model and threshold segmentation and other image processing methods, the calculation is complex and time-consuming, and the application of artificial intelligence and machine learning in image processing is more efficient and accurate. The image processing effect is better. In this paper, the correlative image processing neural network model and its application are analyzed and prospected.

- a) The recognition model of artificial neural network is an artificial intelligence machine learning model that simulates biological nervous system. Using artificial neural network can effectively solve the problem of artificial extraction of image features. Because the artificial extraction of image features requires operators to have rich experience. Based on artificial neural network, image features can be extracted independently, and then image processing, which solves the shortage of artificial feature extraction.
- b) In image processing, in order to extract image features more effectively, artificial neural network needs to set more nodes, which will lead to too many network parameters, resulting in increased learning difficulty or excessive learning affecting the efficiency of image processing.
- c) With the development of artificial intelligence and machine learning, convolutional neural network based on artificial neural network has become an absolute mainstream trend in the application of image processing. Based on convolutional neural network, the algorithm model of image processing is constantly updated, and the capability, speed and quality of image processing are becoming stronger and stronger.
- d) The application of artificial intelligence and machine learning to image processing tends to be more and more lightweight. On the premise of ensuring the accuracy and speed of processing, the main research direction in the future will be to use fewer parameters, reduce the conditions of use and further improve the image processing speed and recognition accuracy.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

PX: Formal Analysis, Project administration, Writing—original draft. JW: Conceptualization, Investigation, Writing—review and editing. YJ: Formal Analysis, Supervision, Writing—review and editing. XG: Writing—review and editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. The study was supported by the National Natural Science Foundation of China (Grant No.: 52378435, 52078065), and the

Science and Technology Innovation Program of Hunan Province (Grant No.: 2023RC3146).

Conflict of interest

Author PX was employed by Zhejiang Infrastructure Construction Group Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Adem, K. (2022). Impact of activation functions and number of layers on detection of exudates using circular Hough transform and convolutional neural networks. *Expert Syst. Appl.* 203, 117583. doi:10.1016/j.eswa.2022.117583
- Alqatawna, A., Rivas Álvarez, A. M., and García-Moreno, S.-C. G. (2021). Comparison of multivariate regression models and artificial neural networks for prediction highway traffic accidents in Spain: a case study. *Transp. Res. procedia* 58, 277–284. doi:10.1016/j.trpro.2021.11.038
- Beni, G., and Wang, J. (1993). "Swarm intelligence in cellular robotic systems," in *Robots and biological systems: towards a new bionics?* (Berlin, Heidelberg: Springer Berlin Heidelberg), 703–712.
- Bera, S., and Shrivastava, V. K. (2020). Effect of pooling strategy on convolutional neural network for classification of hyperspectral remote sensing images. *IET Image Process.* 14 (3), 480–486. doi:10.1049/iet-ipr.2019.0561
- Bhardwaj, T., Mittal, R., Upadhyay, H., and Lagos, L. (2022). "Applications of swarm intelligence and deep learning algorithms for image-based cancer recognition," in *Artificial intelligence in healthcare*, 133–150.
- Brahimi, M., Boukhalfa, K., and Moussaoui, A. (2017). Deep learning for tomato diseases: classification and symptoms visualization. *Appl. Artif. Intell.* 31 (4), 299–315. doi:10.1080/08839514.2017.1315516
- Chauhan, V. K., Dahiya, K., and Sharma, A. (2019). Problem formulations and solvers in linear SVM: a review. *Artif. Intell. Rev.* 52, 803–855. doi:10.1007/s10462-018-9614-6
- Chen, Z.-J., Wu, C. Z., Zhang, Y. S., Huang, Z., Jiang, J. F., Lyu, N. C., et al. (2016). Vehicle behavior learning via sparse reconstruction with ℓ_1 - ℓ_2 minimization and trajectory similarity. *IEEE Trans. Intelligent Transp. Syst.* 18 (2), 236–247. doi:10.1109/tits.2016.2587814
- Cheng, X., Zhang, L., and Zheng, Y. (2018). Deep similarity learning for multimodal medical images. *Comput. Methods Biomechanics Biomed. Eng. Imaging and Vis.* 6 (3), 248–252. doi:10.1080/21681163.2015.1135299
- Dollar, P., Tu, Z., and Belongie, S. (2006). "Supervised learning of edges and object boundaries," in 2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06) (IEEE), 1964–1971.
- Dorigo, M., and Di Caro, G. (1999). "Ant colony optimization: a new meta-heuristic," in Proceedings of the 1999 congress on evolutionary computation-CEC99 (Cat. No. 99TH8406) (IEEE), 1470–1477.
- Dragović, S. (2022). Artificial neural network modeling in environmental radioactivity studies—A review. *Sci. Total Environ.* 847, 157526. doi:10.1016/j.scitotenv.2022.157526
- Dreiseitl, A., and Ohno-Machado, L. (2002). Logistic regression and artificial neural network classification models: a methodology review. *J. Biomed. Inf.* 35 (5–6), 352–359. doi:10.1016/s1532-0464(03)00034-0
- Egrioglu, E., Baş, E., and Chen, M.-Y. (2022). Recurrent dendritic neuron model artificial neural network for time series forecasting. *Inf. Sci.* 607, 572–584. doi:10.1016/j.ins.2022.06.012
- El-Khatib, S. A., Skobtsov, Y. A., and Rodzin, S. I. (2019). Theoretical and experimental evaluation of hybrid ACO-k-means image segmentation algorithm for MRI images using drift-analysis. *Procedia Comput. Sci.* 150, 324–332. doi:10.1016/j.procs.2019.02.059
- Ganji, A., Zhang, M., and Hatzopoulou, M. (2022). Traffic volume prediction using aerial imagery and sparse data from road counts. *Transp. Res. part C Emerg. Technol.* 141, 103739. doi:10.1016/j.trc.2022.103739
- Gao, Y., and Mosalam, K. M. (2018). Deep transfer learning for image-based structural damage recognition. *Computer-Aided Civ. Infrastructure Eng.* 33 (9), 748–768. doi:10.1111/mice.12363
- Gutjahr, W. J. (2000). A graph-based ant system and its convergence. *Future gener. Comput. Syst.* 16 (8), 873–888. doi:10.1016/s0167-739x(00)00044-3
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas, NV, USA, 27–30 June 2016, (IEEE).
- He, K., Zhang, X., Ren, S., and Sun, J. (2014). Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Analysis and Mach. Intell.* 37 (9), 1904–1916. doi:10.1109/tpami.2015.2389824
- He, K., Zhang, X., Ren, S., and Sun, J. (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. pattern analysis Mach. Intell.* 37 (9), 1904–1916. doi:10.1109/tpami.2015.2389824
- Hinton, G. E., Osindero, S., and Teh, Y. W. (2006). A fast learning algorithm for deep belief nets. *Neural Comput.* 18 (7), 1527–1554. doi:10.1162/neco.2006.18.7.1527
- Hinton, G. E., and Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science* 313, 504–507. doi:10.1126/science.1127647
- Howard, A., Sandler, M., Chen, B., Wang, W., Chen, L. C., Tan, M., et al. (2020). "Searching for MobileNetV3," in 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South), 27 October 2019–02 November 2019 (IEEE).
- Huang, H., Lin, L., Tong, R., Hu, H., Qiao, Z., Iwamoto, Y., et al. (2020). "Unet 3+: a full-scale connected unet for medical image segmentation," in ICASSP 2020-2020 IEEE international conference on acoustics, speech and signal processing (ICASSP) (IEEE), 1055–1059.
- Ilie, S., and Bădică, C. (2013). Multi-agent approach to distributed ant colony optimization. *Sci. Comput. Program.* 78 (6), 762–774. doi:10.1016/j.scico.2011.09.001

- Iseensee, F., Jaeger, P. F., Kohl, S. A. A., Petersen, J., and Maier-Hein, K. H. (2021). nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat. methods* 18 (2), 203–211. doi:10.1038/s41592-020-01008-z
- Jin, C., Wang, K., Han, T., Lu, Y., Liu, A., and Liu, D. (2022). Segmentation of ore and waste rocks in borehole images using the multi-module densely connected U-net. *Comput. and Geosciences* 159, 105018. doi:10.1016/j.cageo.2021.105018
- Juan, N. P., and Valdecantos, V. N. (2022). Review of the application of artificial neural networks in ocean engineering. *Ocean. Eng.* 259, 111947. doi:10.1016/j.oceaneng.2022.111947
- Kangling, C. W. F. (2008). “Multilevel thresholding algorithm based on particle swarm optimization for image segmentation,” in Proceedings of the 27th Chinese Control Conference, 16–18.
- Kasabov, N., Scott, N. M., Tu, E., Marks, S., Sengupta, N., Capecci, E., et al. (2016). Evolving spatio-temporal data machines based on the NeuCube neuromorphic framework: design methodology and selected applications. *Neural Netw.* 78, 1–14. doi:10.1016/j.neunet.2015.09.011
- Kennedy, J., and Eberhart, R. (1995). “Particle swarm optimization,” in Proceedings of ICNN’95-international conference on neural networks (IEEE), 1942–1948.
- Khaire, P. A., and Kumar, P. (2021). RGB+ D and deep learning-based real-time detection of suspicious event in Bank-ATMs. *J. Real-Time Image Process.* 18 (5), 1789–1801. doi:10.1007/s11554-021-01155-2
- Khan, A., Vibhute, A. D., Mali, S., and Patil, C. (2022). A systematic review on hyperspectral imaging technology with a machine and deep learning methodology for agricultural applications. *Ecol. Inf.* 69, 101678. doi:10.1016/j.ecoinf.2022.101678
- Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* 25, 84–90. doi:10.1145/3065386
- Kumar, K. S. A., Prasad, A. Y., and Metan, J. (2022). A hybrid deep CNN-Cov-19-Res-Net Transfer learning archtype for an enhanced Brain tumor Detection and Classification scheme in medical image processing. *Biomed. Signal Process. Control* 76, 103631. doi:10.1016/j.bspc.2022.103631
- Lai, S., Jin, L., and Yang, W. (2017). Toward high-performance online HCCR: a CNN approach with DropDistortion, path signature and spatial stochastic max-pooling. *Pattern Recognit. Lett.* 89, 60–66. doi:10.1016/j.patrec.2017.02.011
- Liang, W., Zhang, H., Zhang, G., and Cao, H. x. (2019). Rice blast disease recognition using a deep convolutional neural network. *Sci. Rep.* 9 (1), 2869–2910. doi:10.1038/s41598-019-38966-0
- Lin, H., Li, L., Wang, H., Wang, Y., and Ma, Z. (2020). Traffic flow prediction using SPGAPSO-CKRVM model. *Rev. d’Intelligence Artif.* 34 (3), 257–265. doi:10.12820/ria.340303
- Long, J., Shelhamer, E., and Darrell, T. (2015). “Fully convolutional networks for semantic segmentation,” in Proceedings of the IEEE conference on computer vision and pattern recognition.
- Lorente, D., Zude, M., Regen, C., Palou, L., Gómez-Sanchis, J., and Blasco, J. (2013). Early decay detection in citrus fruit using laser-light backscattering imaging. *Postharvest Biol. Technol.* 86, 424–430. doi:10.1016/j.postharvbio.2013.07.021
- Luchini, C., Pea, A., and Scarpa, A. (2022). Artificial intelligence in oncology: current applications and future perspectives. *Br. J. Cancer* 126 (1), 4–9. doi:10.1038/s41416-021-01633-1
- Lyu, T., Zhao, W., Zhu, Y., Wu, Z., Zhang, Y., Chen, Y., et al. (2021). Estimating dual-energy CT imaging from single-energy CT data with material decomposition convolutional neural network. *Med. image Anal.* 70, 102001. doi:10.1016/j.media.2021.102001
- Marini, F., and Walczak, B. (2015). Particle swarm optimization (PSO). A tutorial. *Chemom. Intelligent Laboratory Syst.* 149, 153–165. doi:10.1016/j.chemolab.2015.08.020
- Medina-Salgado, B., Sánchez-DelaCruz, E., Pozos-Parra, P., and Sierra, J. E. (2022). Urban traffic flow prediction techniques: a review. *Sustain. Comput. Inf. Syst.* 35, 100739. doi:10.1016/j.suscom.2022.100739
- Min, W., Li, X., Wang, Q., Zeng, Q., and Liao, Y. (2019). New approach to vehicle license plate location based on new model YOLO-L and plate pre-identification. *IET Image Process.* 13 (7), 1041–1049. doi:10.1049/iet-ipr.2018.6449
- Niu, X.-X., and Suen, C. Y. (2012). A novel hybrid CNN-SVM classifier for recognizing handwritten digits. *Pattern Recognit.* 45 (4), 1318–1325. doi:10.1016/j.patcog.2011.09.021
- Pontil, M., and Verri, A. (1998). Support vector machines for 3D object recognition. *IEEE Trans. Pattern Analysis and Mach. Intell.* 20 (6), 637–646. doi:10.1109/34.683777
- Qin, J., Shen, X., Mei, F., and Fang, Z. (2019). An Otsu multi-thresholds segmentation algorithm based on improved ACO. *J. Supercomput.* 75, 955–967. doi:10.1007/s11227-018-2622-0
- Ragul, K., and Cloudin, S. (2021). “Identification of criminal and non-criminal face using deep learning and image processing,” in 2021 International Conference on System, Computation, Automation and Networking (ICSCAN) (IEEE).
- Rahat, A. M., Kahir, A., and Masum, A. K. M. (2019). “Comparison of Naive Bayes and SVM Algorithm based on sentiment analysis using review dataset,” in 2019 8th International Conference System Modeling and Advancement in Research Trends (SMART) (IEEE).
- Ronneberger, O., Fischer, P., and Brox, T. (2015). “U-net: convolutional networks for biomedical image segmentation,” in International Conference on Medical Image Computing and Computer-Assisted Intervention (Springer International Publishing).
- Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back propagating errors. *Nature* 323 (6088), 533–536. doi:10.1038/323533a0
- Sahu, S. K., and Pandey, M. (2023). An optimal hybrid multiclass SVM for plant leaf disease detection using spatial Fuzzy C-Means model. *Expert Syst. Appl.* 214, 118989. doi:10.1016/j.eswa.2022.118989
- Sen Sarma, S., and Bandyopadhyay, S. K. (1989). Some sequential graph colouring algorithms. *Int. J. Electron. Theor. Exp.* 67 (2), 187–199. doi:10.1080/00207218908921070
- Schapiro, R. E. (1999). A brief introduction to boosting. *Proc. IJCAI* 99 (999), 1401–1406.
- Singh, A., Singh, K. K., and Saravana, V. (2021). Real-time intelligent image processing for security applications. *J. Real-Time Image Process.* 18, 1787–1788. doi:10.1007/s11554-021-01169-w
- Smith, S. D., Amos, J. D., Beck, K. N., Colvin, L. M., Franke, K. S., Liebl, B. E., et al. (2014). Refinement of a protocol for the induction of lactation in nonpregnant nonhuman primates by using exogenous hormone treatment. *J. Am. Assoc. Laboratory Animal Sci.* 53 (6), 700–707.
- Sun, Y., Chen, Y., Wang, X., and Tang, X. (2014b). “Deep learning face representation by joint identification-verification,” in *Advances in neural information processing systems*, 27.
- Sun, Y., Wang, X., and Tang, X. (2014a). “Deep learning face representation from predicting 10,000 classes,” in Proceedings of the IEEE conference on computer vision and pattern recognition (IEEE), 1891–1898.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., et al. (2014). *Going deeper with convolutions*. IEEE Computer Society.
- Szegedy, C., Ioffe, S., Vanhoucke, V., and Alemi, A. (2017). Inception-v4, inception-resnet and the impact of residual connections on learning. *Proc. AAAI Conf. Artif. Intell.* 31 (1). doi:10.1609/aaai.v31i1.11231
- Tang, J., Liu, G., and Pan, Q. (2021). A review on representative swarm intelligence algorithms for solving optimization problems: applications and trends. *J. Automatica Sinica* 8 (10), 1627–1643. doi:10.1109/jas.2021.1004129
- Tang, M., and Chen, F. (2013). Facial expression recognition and its application based on curvelet transform and PSO-SVM. *Optik* 124 (22), 5401–5406. doi:10.1016/j.ijleo.2013.03.116
- Trinh, D. H., Luong, M., Rocchisani, J. M., Pham, C. D., and Dibos, F. (2011). “Adaptive medical image denoising using support vector regression,” in Computer Analysis of Images and Patterns: 14th International Conference, CAIP 2011, Seville, Spain, August 29–31, 2011 (Berlin Heidelberg: Springer), 494–502.
- Turing, A. M. (1950). Computing machinery and intelligence. *Mind* 59, 433–460.
- Ulku, I., and Akagunduz, E. (2019). *A survey on deep learning-based architectures for semantic segmentation on 2d images*. arxiv preprint arxiv:1912.10230.
- Wu, X., Zhao, J., and Shen, M. (2010). Face processing in human-computer interaction using curvelet analysis. *J. Image Graph.* 15, 1309–1317.
- Wu, Y., and Feng, J. (2018). Development and application of artificial neural network. *Wirel. Personal. Commun.* 102, 1645–1656. doi:10.1007/s11277-017-5224-x
- Yang, Y., Zhong, Z., Shen, T., and Lin, Z. (2018). “Convolutional neural networks with alternately updated clique,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* 15, 2413–2422. doi:10.1109/cvpr.2018.00256
- Yang, Z., Zhang, W., and Feng, J. (2022). Predicting multiple types of traffic accident severity with explanations: a multi-task deep learning framework. *Saf. Sci.* 146, 105522. doi:10.1016/j.ssci.2021.105522
- Yang, Z. R. (2006). A novel radial basis function neural network for discriminant analysis. *IEEE Trans. neural Netw.* 17 (3), 604–612. doi:10.1109/tnn.2006.873282
- Yuan, Q., Wei, Z., Guan, X., Jiang, M., Wang, S., Zhang, S., et al. (2019). Toxicity prediction method based on multi-channel convolutional neural network. *Molecules* 24 (18), 3383. doi:10.3390/molecules24183383
- Yusri, I. M., Abdul Majeed, A., Mamat, R., Ghazali, M., Awad, O. I., and Azmi, W. (2018). A review on the application of response surface method and artificial neural network in engine performance and exhaust emissions characteristics in alternative fuel. *Renew. Sustain. Energy Rev.* 90, 665–686. doi:10.1016/j.rser.2018.03.095

Zhang, L., and Lim, C. P. (2020). Intelligent optic disc segmentation using improved particle swarm optimization and evolving ensemble models. *Appl. Soft Comput.* 92, 106328. doi:10.1016/j.asoc.2020.106328

Zhang, X., and Dahu, W. (2019). Application of artificial intelligence algorithms in image processing. *J. Vis. Commun. Image Represent.* 61, 42–49. doi:10.1016/j.jvcir.2019.03.004

Zhang, Y., Hua, Q., Wang, H., Ji, Z., and Wang, Y. (2023). Gaussian-type activation function with learnable parameters in complex-valued convolutional neural network and its application for PolSAR classification. *Neurocomputing* 518, 95–110. doi:10.1016/j.neucom.2022.10.082

Zhou, D. X. (2018). Deep distributed convolutional neural networks: universality. *Analysis Appl.* 16, 895–919. doi:10.1142/s0219530518500124

Zhou, Y., Ren, Y., Xu, E., Liu, S., and Zhou, L. (2022). Supervised semantic segmentation based on deep learning: a survey. *Multimedia Tools Appl.* 81, 29283–29304. doi:10.1007/s11042-022-12842-y

Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., and Liang, J. (2018). “UNet++: a nested U-net architecture for medical image segmentation,” in *Deep learning in medical image analysis and multimodal learning for clinical decision support*. Editors D. Stoyanov, Z. Taylor, G. Carneiro, T. Syeda-Mahmood, A. Martel, L. Maier-Hein, et al. (Cham: Springer International Publishing), 3–11.