International Journal of Science and Engineering Applications
Volume 14-Issue 04, 35 – 41, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1404.1006

Canine Image Recognition Classification Based on Improved DenseNet121 Model

Bi Jilin
School of Electronic Information and Electrical Engineering
Yangtze University
Jingzhou, China

Abstract: Recognition of animal dog species has always been the focus of the image recognition field, in order to better recognize the canine species in the image and help the society to manage the family pets, this paper discusses to propose a model based on the combination of the YOLOv8 recognition algorithm and the improvement of the network structure of the DenseNet121 for the recognition of canine species, through the addition of the YOLOv8 algorithm to the front of DenseNet121 and the addition of the attention module (CBAM) inside each Dense Block to solve the problems of gradient disappearance, parameter redundancy, and insufficient feature propagation in traditional convolutional network by unique connection. algorithm in front of the DenseNet121 and adding the Attention Module (CBAM) inside each Dense Block, the problems of gradient vanishing, parameter redundancy, and insufficient feature propagation in traditional convolutional networks are solved by a unique connection, which can more accurately recognize and classify the dog in the image. This experiment uses an image dataset containing 120 dog breeds from all over the world from the Kaggle website for image recognition. The cutting-edge deep learning framework pytorch and computationally powerful GPUs were selected to use deep neural networks to train and test the network on dog images, which ultimately improved the accuracy and robustness of the model on image classification and confirmed the reliability of the model.

Keywords: image recognition; deep learning; convolutional neural network; DenseNet121

1. INTRODUCTION

Animal species recognition has been one of the research hotspots in the field of computer vision and artificial intelligence. Effective solutions are urgently needed for both rare animal protection and daily pet management. In recent years, the problem of stray pets has become increasingly serious, bringing many challenges to society. Many people give up adopting pets for various reasons, leading to the increasing number of stray dogs. To address this problem, this study proposes an improved DenseNet121 convolutional neural network for efficient recognition and classification of canine images to aid in community and household pet management.

Convolutional neural networks (CNNs) have significant advantages in the field of image processing, the core of which lies in the extraction of image features by means of convolutional layers[1] , which is significantly different from traditional fully connected neural networks. In order to cope with the overfitting problem caused by the large amount of model parameters, in this study, the convolutional and pooling layers are used to extract features from the input image, and the extracted features are subsequently input into the fully connected network for classification[2].

During the development of convolutional neural networks, many classical network structure models have emerged, such as LeNet, AlexNet, VGGNet, and InceptionNet, which play an important role in the field of image recognition[3]. The core idea of DenseNet is the dense connectivity: the input of each convolutional layer includes not only the output of the previous layer, but also the outputs of all the preceding layers' outputs. However, as the number of network layers increases, degradation of the model performance occurs.[4] Therefore, introducing an attention mechanism into the DenseNet121 model is an effective improvement method to help the model better focus on key feature regions in the image, thus improving the classification performance.

2. ANALYZING NETWORK MODELS

2.1 DenseNet121 model

DenseNet121 is a densely connected convolutional neural network based on a unique connectivity and efficient feature utilization mechanism.[5]In each Dense Block, the input of each layer not only includes the output of the previous layer, but also integrates the outputs of all previous layers. This densely connected design greatly facilitates feature reuse and efficient flow of information, effectively mitigating the gradient vanishing problem and enabling the network to propagate gradients more

International Journal of Science and Engineering Applications
Volume 14-Issue 04, 35 – 41, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1404.1006

efficiently, thus making it easier to train. At the same time, because features are utilized multiple times in the network, DenseNet121 excels in parameter efficiency, typically requiring fewer parameters to achieve the same or even better performance than other deep network models such as VGG or ResNet. In addition, the ability of lower layer features to propagate directly to the output layer not only strengthens the utilization of features, but also allows the network to better capture and utilize information from earlier extracted features.[6] These properties of DenseNet121 allow for a certain level of overfitting suppression when dealing with small datasets, as well as support for the construction of very deep network structures without training difficulties due to the increased depth. Despite the large number of connections in the network, the fact that feature maps rather than parameters or gradients are passed between layers, as well as the fact that DenseNet121 performs well in terms of computational and memory efficiency, make it highly practical and efficient in real-world applications.The structure of the DenseNet121 network is shown in Figure. 1.

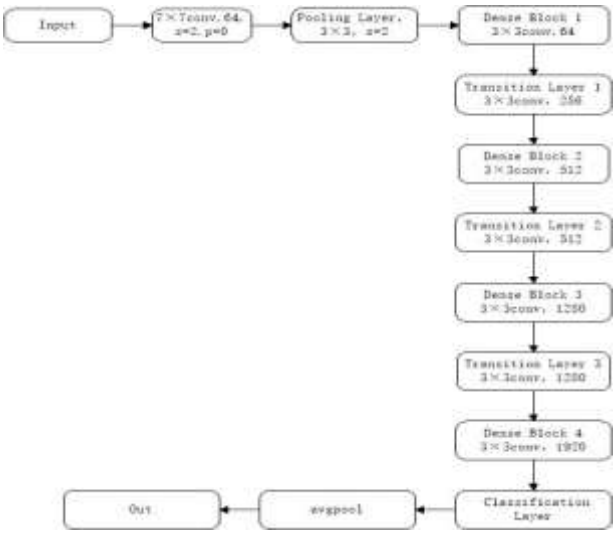

Figure .1 DenseNet121 network structure

2.2 Channel Attention Mechanisms

Attention mechanism is a key idea in the field of deep learning and computer vision, and has been widely used in recent years in many fields such as natural language processing and image recognition.[7] Its core role is to focus on the more critical information in the current task, so as to extract more detailed features, while reducing the attention of other irrelevant information and suppressing feature extraction that is not beneficial to the task. There are various types of attention mechanisms, among which the more commonly used are channel attention and spatial attention, such as SE modules and CBAM modules[8] .

SE Attention mechanism (Squeeze-and-Excitation Networks) is a typical channel attention mechanism. The core of the mechanism is to assign an attention weight to each channel in the feature graph, so that the network pays more attention to the feature channels that are useful for the current task, while suppressing the channels that are not very helpful for the task. This mechanism is realized through the steps of "Squeeze" and "Excitation", which can dynamically adjust the importance of each channel and enhance the model's ability to perceive key features. In this study, we choose to adopt the SE module to enhance the model's ability to perceive and process key information. In this paper, the channel attention mechanism is introduced into the DenseNet121 network to optimize the network, so that the network can fully access the effective information in the features and increase the perceptual range of the convolutional kernel.[9]The structure of the channel attention layer is shown in Figure. 2.

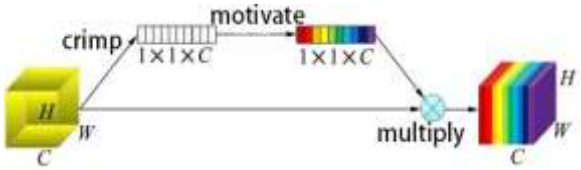

Figure. 2 Structure of channel attention

The core operation of the mechanism consists of two steps: Squeeze and Excitation. In the "Squeeze" step, the model is pooled by the Global Average Pooling operation.

$$Z_c = \frac{1}{H \times W} \sum_{i=1}^{H} \quad \sum_{j=1}^{W} X_{c,\ i,\ j} \qquad (1)$$

Where: Zc denotes the output after global pooling, subscript c denotes the channel number; H, W denote the height and width of the feature matrix in each channel; Xc(i, j) denotes the i,jth element of the feature matrix of the cth channel. The feature maps of each channel are processed to compress the spatial dimensions to a single value in order to obtain the global information of each channel. The purpose of this step is to aggregate the characteristic responses of each channel to form a statistic that describes the importance of the channel[10] .

Subsequently, in the "excitation" step, the model nonlinearly transforms the compressed channel descriptors through a Fully

International Journal of Science and Engineering Applications
Volume 14-Issue 04, 35 – 41, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1404.1006

Connected Layer to learn the relative importance of each channel.

$$n = F_{ex}(y, W) = \sigma(g(y, W)) = \sigma(W2\delta(W1y)) \quad (2)$$

where: n is the channel weight vector after the excitation operation; Fex(-) denotes the excitation operation on the channel information; y denotes the output of the squeeze-compress operation; W denotes the weight matrix of the fully connected layer; σ(-) denotes the sigmiod activation function, which is used to normalize the channel weights to be between 0 and 1; g(-) denotes the gating of y and W; δ(-) denotes the RELU activation function, which is used to introduce nonlinear characteristics. used to introduce nonlinear properties:W1,W2 denote the weight matrices of the 2 fully connected layers.

This typically involves two fully connected layers[11] with a ReLU activation function sandwiched between them to introduce nonlinear properties and facilitate feature selection. Ultimately, the learned importance is converted into channel weights by means of a Sigmoid function, which reflect the degree of contribution of each channel to the task at hand.

$$XC = Sc \times X\ c \quad (3)$$

Where: XC denotes the output after the squeeze excitation o peration;Sc, Xc denotes the output of the squeeze-compression operation with excitation operation for the cth channel.

In this way, the channel attention mechanism is able to automatically learn and reinforce those feature channels that are most helpful for the task at hand, while suppressing those that are less important, thus allowing the neural network to focus more on useful feature representations, improving the performance and generalization of the model[12].

2.3 YOLOv8 image classification algorithm

The YOLOv8 network consists of three modules, Backbone, Neck and Head.[13] Among them, the Backbone module is responsible for downsampling operation on the input image, extracting the multilevel features of the image through the network structure of different depths, and finally generating the feature map of the image. On this basis, this paper extracts the Backbone module of YOLOv8 separately and uses it as a feature extraction module in order to extract representative features. The specific architecture is shown in Figure 3.

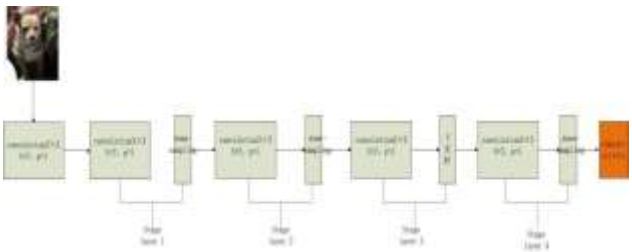

Figure. 3 YOLOv8 classification feature network architecture

From Figure. 3 we can see that the Backbone module of YOLOv8 progressively downsamples and extracts features from the input image through four stages.[14]Each stage contains multiple convolutional layers, where the convolutional layer with a step size of 2 is used to reduce the spatial size of the feature map and the convolutional layer with a step size of 1 is used to further extract features. Specifically, the first stage downsamples the size of the input image from 224 × 224 to 112 × 112, the second stage further downsamples it to 56 × 56, the third stage to 28 × 28, and finally the fourth stage to 14 × 14. The number of channels in the convolutional layers of each stage is gradually increased from 64 to 512 to capture richer feature information. This multi-stage downsampling and feature extraction design enables the Backbone module to efficiently capture multi-scale features from local to global, providing a robust feature representation for subsequent target detection and classification tasks.

2.4 Improved DenseNet121 network model ing

In order to be able to recognize images more accurately and efficiently, this thesis introduces the attention mechanism and CAM module as well as the YOLOv8 image classification algorithm on the basis of the DenseNet121 network model, and proposes a brand new structure of the DenseNet121 network model. In a densely connected network, each layer is connected to all previous layers. This design makes the flow of information in the network more efficient and helps the propagation of gradients and reuse of features. Each dense block consists of three parts: batch normalization (BN), ReLU activation function and 3D convolutional layer (i.e., 3x3x3 convolutional kernel). Batch Normalization helps to reduce the internal covariate bias, which not only speeds up the training but also improves the accuracy of the model.The ReLU activation function helps to reduce the probability of gradient vanishing.

Between each dense block, a transition layer is set up, which consists of a batch normalization, a 1x1 convolutional layer (to maintain the original size of the feature map), and a maximum

International Journal of Science and Engineering Applications
Volume 14-Issue 04, 35 – 41, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1404.1006

pooling layer to reduce the feature dimensionality.Each layer of DenseNet is connected to all the networks in the previous layer, which allows the network to directly access the input features and reuse features with better robustness through the connections on the channels.

In addition, an attention mechanism is added to the densely connected structure, as shown in Figure 4. This attention me chanism can further enhance the performance of the model b y focusing on the feature map channels that are more helpful for the task at hand and suppressing those that are less usef ul. In this way, the model can focus more on the features th at are critical to the task, thus improving the accuracy and ef ficiency of the recognition.

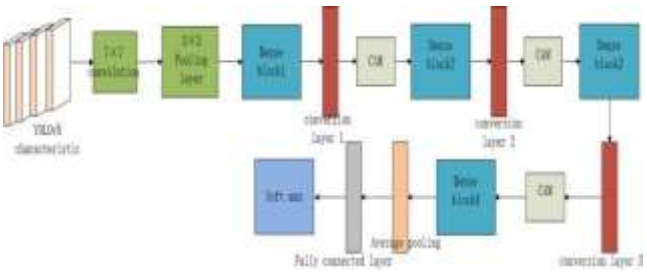

Figure. 4 DenseNet121 model with the addition of channel attentio n mechanism

3. EXPERIMENTS AND EXPERIMENT AL RESULTS

3.1 Experimental data set

For the research topic canine image recognition based on improved DenseNet121 network model for classification prediction of dogs in the dataset. The dataset used for the experiments is the Stanford Dogs dataset from the kaggle competition, a large canine recognition dataset created by the Stanford University Computer Vision Laboratory. This dataset contains 120 different breeds of dogs with approximately 100 images of each breed, totaling 20,580 images of dogs.

3.2 Data pre-processing

The dataset is stored in a folder named CATPNG, ABtrain folder, ABtest folder, ABtest_label.txt and ABtrain_label.txt are created under CATPNG folder, and the training set and test set are put into the corresponding folders, and the feature labels of the training set and test set are stored in the corresponding text files. corresponding text files. The images in the dataset are uniformly resized to 224*224*3 dimensions before being sent to the model for training, and no data enhancement is performed because the data in the kaggle dataset is relatively clean and the dataset is large.

3.3 Data set partitioning

Firstly, we integrate the dataset for classification and categorize the different kinds of dogs in the data by data labels as in Figure. 5 In this experiment, the number of training set and test set is divided in the ratio of 4:1, i.e., out of 20580 images, there are 16464 training set and 4116 test set respectively, so that the training set and the test set will never intersect each other, and the number of different categories in the training set and test set is the of canine images are the same.

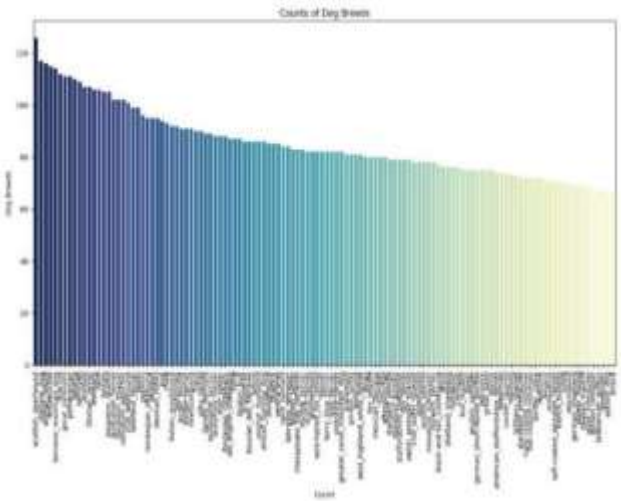

Figure .5 Classification of dog breed pictures

3.4 Experimental results

In this experiment, we used the improved DenseNet121 network model for training. The experimental environment is configured as follows: the processor is an Intel (Core) i7-9750H CPU at 2.50 GHz, and the graphics card is a GTX 1650 with 16 GB of memory size, using the parallel computing architecture CUDA 10.2. The software environment consists of Python version 3.8 and Keras as a deep learning framework[15].

Since the model structure is more complex after improvement and the amount of data to be processed is larger, we set the number of samples per training (batch_size) to a smaller 16 and the training iteration period (epoch) to 15. During the training process of configuring the DenseNet121 network, we chose the Adam optimizer with faster convergence[16] , the loss function was chosen to be the cross-entropy loss function, and L2 regularization was added to mitigate the overfitting problem o

International Journal of Science and Engineering Applications
Volume 14-Issue 04, 35 – 41, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1404.1006

Table 1 Comparative experimental results

MOULD	Improved DenseNet121	DenseNet121	ResNet18	inception10	vgg16	alexnet8
TEST SET ACCURACY	0.9941	0.8466	0.7212	0.7634	0.6431	0.6845
LOSS FUNCTION	0.1324	0.4536	0.4581	0.4231	0.8132	0.7312

the model.[17] At the end of the training, the model achieved an accuracy of 99.4% on the training set and 95.2% on the validation set. Figure 6 illustrates the change in accuracy (Accuracy) as well as loss function value (Loss) for the training and validation sets while the model is running. From the figure, it can be seen that the overfitting phenomenon was successfully avoided during the training process.

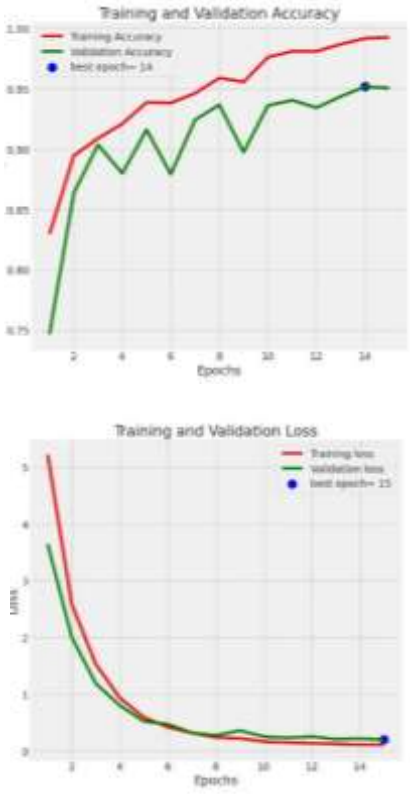

Figure. 6 Improved DenseNet121 network model training results

3.5 Comparative test results

In order to demonstrate the improvement of the improved DenseNet121 network model over the original DenseNet121 as well as other common models (DenseNet and ResNet18),[18] comparative experiments were conducted. The experiments were conducted on the same equipment as the improved DenseNet121 model was trained with to ensure comparable results. The evaluation criteria were the average loss value and accuracy of the last four iterations on the test set.[19]

The comparison results, as shown in Table 1, show that the recognition accuracy of the improved DenseNet121 model exceeds that of the original DenseNet121 model on the dog dataset, thus confirming the validity and usefulness of our proposed model.

3.6 Individual identification and visual an alysis of dog breed classification

In order to further analyze and compare the classification effect of the improved DenseNet121 model, three canine images were randomly selected from the dog classification number test set and plotted to compare the results of each classification, as shown in Fig. 7 and Fig. 8 Fig. 9.

Figure. 7 malinois dog classification prediction Figure 8 pembroke dog classification prediction Figure 9 redbone dog classification pr ediction

From the pictures we can see that we used the pictures in our test dataset for the model to predict them, and the results show that our improved DenseNet121 model is able to classify the canines in the pictures more accurately. Among them, Figure 7 shows malinois dogs with 63.66% accuracy, Figure 8 shows pembroke dogs with up to 85.36% accuracy, and Figure 9 shows redbone dogs with 78.87% accuracy.

In order to verify that the improved model can have accurate classification performance, we randomly select a number of pictures of 8 canine breeds from the dataset and test the classification results and draw the confusion matrix as shown in Fig. 10.

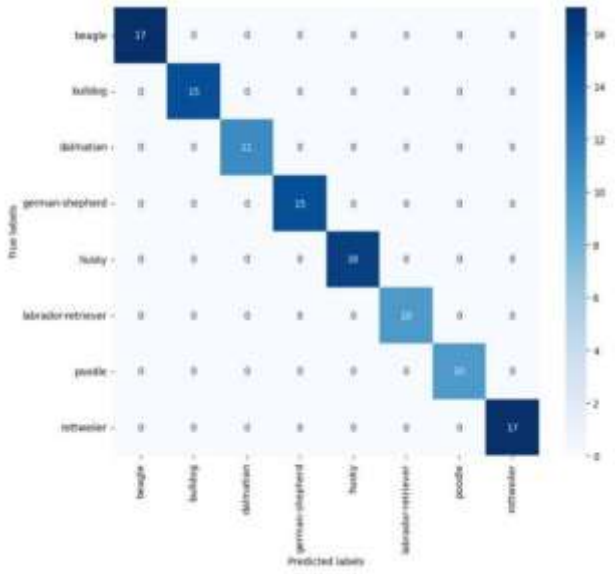

Figure .10 Classification Confusion Matrix

From this confusion matrix we can see that it can perfectly classify the dataset that we have given to it, probably because we have given to it the dataset that has been trained and will have a perfect classification strategy.

From the results, it can be seen that on the dog dataset, the improved DenseNet121 network model of this thesis has a higher recognition accuracy, which verifies the feasibility and effectiveness of the model of this thesis.

4. CONCLUSION

In order to improve the accuracy of canine image recognition, this study improves the DenseNet121 network model and proposes a network structure based on the improved DenseNet121 network. The improved model adds the YOLOv8 image classification algorithm in front of the convolutional layer of the original network and introduces the CBAM attention mechanism module. On the basis of retaining the original structure, these improvements not only effectively reduce the degradation phenomenon and prevent overfitting, but also realize the enhancement of local features and suppress unnecessary feature extraction, thus improving the accuracy of image recognition. The feasibility and effectiveness of the network model are demonstrated through comparative experimental validation. Compared with the original network model, the improved model has a significant improvement in recognition accuracy and can more accurately recognize categories in images.

5. REFERENCES

[1] Zhang Yuhong, Bai Lianxiang, Meng Fanjun, et al. Convolutional neural network in image recognition [J]. New Technology and New Process, 2021, 397(1):52-55.

[2] Cui Yongyi, Qu Fang. Experimental Discussion on Fire Image Recognition Based on Deep Learning［J］. Journal of Physics: Conference Series, 2021, 2066(1).

[3] Ke Zhang, Xiaohan Feng, Yurong Guo, et al. A review of deep convolutional neural network models for image classification[J]. A review of deep convolutional neural network models for image classification［J］. Chinese Journal of Image and Graphics, 2021, 26（10）:2305-2325.

[4] Hang G, Gi Zu, Van Der Maaten L. Densely Connected Convolutional Networks ［J］.Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 4700-4708

[5] Xin Wang,Wenjing Zhang,Wei Shi,et al. Research on li ghtweight modeling of knowledge migration DenseNet post-training pruning for social insect recognition[J]. Jo urnal of Ningxia University(Natural Science Edition),20 24,45(03):307-314.DOI:10.20176/j.cnki.nxdz.000055.

[6] Ma Yongjie, Liu Peipei. Convolutional neural network image classification algorithm based on DenseNet evolution [J]. Advances in Laser and Optoelectronics, 2020, 57 (24) :50-57.

[7] Li Jx, Sun J, Li C, et al. Joint diagnosis and segmentat ion of neocoronary arthritis by integrating multiple atte ntion mechanisms［J］. Chinese Journal of Image Grap

hics, 2022, 27(12): 3651-3662.

[8] Zhu Lei, Tong Chao, Dong Liang, et al. Gait recognition algorithm based on residual network and attention mechanism[J]. Telecommunications Technology, 2022, 62(12):1723-1728

[9] Jauhari K ,Rahman Z A ,Huda A M , et al. A hybrid deep learning-based approach for on-line chatter detection in milling using deep stem-inception networks and residual channel-spatial attention mechanisms[J]. Processing,2025,226112357-112357.

[10] Wang Radius, Yan She-Feng, Mao Linlin, et al. Hydroacoustic target recognition network based on channel grouping attention mechanism[J]. Signal Processing,2025,41(03):524-532.

[11] Sahragard E ,Farsi H ,Mohamadzadeh S .Advancing semantic segmentation: enhanced UNet algorithm with attention mechanism and deformable convolution.[J].PloS one,2025,20(1):e0305561.

[12] Yang Kefan,Wei Xinkai,Ma Yan,et al.Attention-YOLOv5s: a YOLOv5s algorithm introducing attention mechanism[J]. Modern Information Technology,2025,9(05):25-32+38.DOI:10.19850/j.cnki.2096-4706.2025.05.005.

[13] Dillon Reis, Jordan Kupec, Jacqueline Hong, et al. (2023). Real-Time Flying Object Detection with YOLOv8. arXiv, abs/2305.09972

[14] Jiang Xiangkui, Lu Qi, Dong Chao, et al. A diabetic retinopathy detection algorithm based on YOLOv8n[J/OL]. Journal of Xi'an University of Posts and Telecommunications,1-9[2025-04-07].http://kns.cnki.net/kcms/detail/61.1493.TN.20250306.1700.004.html.

[15] Shu Jun Liu, Shengyu Wu. Research on insect image recognition based on convolutional neural network and Tensorflow [J]. Light Source and Lighting, 2022(4):70-73.

[16] Dangi S ,Kumar D ,Khurana V .BAAO: Bayesian and Adam optimizer for fault prediction in self-driving software systems using deep learning-based hyperparameter tuning[J].International Journal of Information Technology,2024,17(2):1-10.

[17] LI Hong .Research on image feature extraction and classification based on deep learning[J].Academic Journal of Computing & Information Science, 2025,8(1)

[18] Wei Yi, Chen LP. Attention mechanism-based attitude estimation technique for deep learning sports [J]. Electronic Design Engineering, 2023, 31(2):152-155.

[19] Zhang Y ,Ning C ,Yang W .An automatic cervical cell classification model based on improved DenseNet121. [J].Scientific reports,2025,15(1):3240.