Multimodal Application of GAN in the Image Recognition of Wheat Diseases and Insect Pests

Bing Li*, Shaoqing Yang, Zeqiang Wang
College of Modern Information Technology, Henan Polytechnic, ZhengZhou 450046, China

***Abstract*—"Food is the most important thing for the people", Food is intricately linked to both the national economy and the livelihood of the people, serving as a vital material for our daily existence. Wheat, standing as one of the three core grain crops, holds paramount importance in safeguarding national food security. However, the wheat planting process remains constantly exposed to a diverse array of environmental factors, ranging from the intensity of light to fluctuations in temperature, soil fertility, fertilizer application methods, and water availability. Occasionally, these variables trigger diseases and insect infestations that can seriously affect wheat yield and quality if not promptly and effectively addressed. Therefore, it is imperative to manage these challenges in a timely and effective manner, ensuring the safety and integrity of wheat production, which in turn guarantees the stability of our national food supply. Traditional methods of manual detection of pests and diseases mainly rely on naked eye observation and manual statistics. Such solutions are highly subjective, have low timeliness, and difficult to unify precision. With the development of computer technology and deep learning, more and more research and applications have been carried out to address the shortcomings of traditional manual detection methods. In this study, deep learning is combined with the application of disease and insect pest recognition. Studying wheat powdery mildew, scab, leaf rust, and midge, convolutional and capsule networks are investigated for pest recognition, establishing an image recognition system for wheat diseases and pests.**

***Keywords*—Deep Learning; Identification of diseases and insect pests; Image classification; System development**

I. Introduction

Wheat, a major food crop, faces challenges from diseases and insect pests triggered by environmental factors [1, 2]. Prompt and accurate identification is crucial to prevent production losses and potential crop failure. Rust, a common menace to wheat crops, can wreak havoc on yields. In epidemic years, it can reduce production by a substantial 20% to 30%. And in extreme cases, the damage can be even more devastating, exceeding 50% and threatening the very existence of wheat production [3, 4]. The figures from the Shandong Plant Protection Research Institute are particularly startling. From 2000 to 2018, the losses attributed to diseases and insect pests in China's prime wheat-growing regions amounted to a staggering 17.67 million tons. That's a loss equivalent to the food supply of nearly 289 million people.

The prevention and prompt diagnosis of wheat diseases and insect pests are imperative for minimizing their detrimental effects on production, yet the unpredictable nature of the agricultural environment poses significant obstacles in the prevention of such threats. Therefore, timely diagnosis and treatment become paramount. Traditionally, disease and pest detection has relied on manual methods, involving naked-eye judgments and manual statistics [5, 6]. The automatic feature extraction function of deep learning enables the automatic classification and recognition of wheat pest images by learning the inherent patterns and characteristics of sample data. This overcomes the limitations of manual recognition in terms of timeliness, subjectivity, and potential damage, offering a novel scientific approach to wheat pest recognition.

A multi-channel network model, CNN-Caps Nets, is established based on convolutional and capsule networks, consisting of multiple conv, pooling, primary capsule, and SoftMax layers. The convolution kernel transmitted by the convolution layer is received by the primary capsule layer, and more image features are extracted for image classification. By comparison, the CNN-Caps Nets model has the best classification effect under the structure of four channels and the number of capsules in the capsule layer is 16. The recognition accuracy of wheat powdery mildew, scab, leaf rust and midge images are 90%, 71%, 91% and 58.3%, respectively. A comprehensive image-sharing database for wheat pests and diseases was established, and a corresponding image recognition system was designed and developed, leveraging the CNN-CapsNet model for effective image classification.

II. Materials and Methods

A. Data Acquisition and Data Set Construction

The quantity and quality of wheat disease and insect pest image samples will directly affect the efficiency and accuracy of subsequent image segmentation and image classification. Due to environmental and regional factors, it is difficult to collect images of wheat diseases and insect pests. This study obtains images with high quality, obvious features, and easy recognition from public data sets (LWDCD, Wheat Leaf Dataset, CGIAR, IDADP, IP102), agricultural databases (National Agricultural Science Data Center, Agricultural Big Data, etc.) and Baidu Gallery the image data is used as the research object, as shown in Table I.

Data sets are essential for training pest classification models. Obtain image data from Wheat-ORL shared database, classify them and collect them in different folders. Using Python, read all pic files in a folder, rename and categorize diseases/insects, then record names and labels in a CSV file as a dataset. The specific data format is shown in Table II.

TABLE I. DATA SOURCES

Wheat Pests and Diseases Image Categories	Number of pictures			Total
	Dataset	Agricultural Databases	Baidu Gallery	
Pest-free	245	0	55	300
Wheat powdery mildew	470	10	20	500
Wheat scab	157	3	30	190
Wheat leaf rust	445	5	100	550
Wheat midge	30	10	20	60

TABLE II. DATASET DATA FORMATS

Image Name	Category	Memo
Heal _ 0. jpg	health	Pest-free
Bfb _ 0. jpg	bfb	Wheat powdery mildew
Cmb _ 0. jpg	cmb	Wheat scab
Yxb _ 0. jpg	yxb	Wheat leaf rust
Xjc _ 0. jpg	xjc	Wheat midge

The 1600-image dataset is divided into training and test sets at a 8:2 ratio for machine learning requirements. The distribution details are shown in Table III.

TABLE III. DATA SET DISTRIBUTION

Categories	Date set	
	Training Set	Testing Set
Pest-free	240	60
Wheat powdery mildew	400	100
Wheat scab	152	38
Wheat leaf rust	440	110
Wheat midge	48	12

B. Graphic Pre-processing

In the process of image generation, it will be affected by noise, insufficient or excessive illumination, inappropriate shooting angle, etc., resulting in a decrease in image quality. In order to improve the accuracy of image feature extraction segmentation and classification smoothing filtering and sharpening of the image can better distinguish the target disease spots pest areas and background in the image.

Smoothing filtering is a low-frequency spatial domain filtering tech to eliminate noise [7]. For different noise characteristics, selecting the corresponding filtering technology can achieve very obvious results. In OpenCV processing library, two kinds of filters are commonly used: Gaussian filter and bilateral filter.

The Gaussian filter is a linear filtering technique that finds extensive application in image smoothing and blurring [8]. When it comes to digital image processing, Gaussian noise is a commonly encountered type of noise. Therefore, Gaussian filtering is extensively utilized in images that are affected by this type of noise. Its basic principle is to achieve image smoothing by weighted averaging the values of each pixel in the image itself and other pixels in the neighborhood. The two-dimensional Gaussian function is the basis for building a Gaussian filter, and the function formula is shown in Eq. (1):

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}} \quad (1)$$

Bilateral filtering is a nonlinear filter that can simultaneously reduce noise, smooth images and save edges. The filter consists of two functions: two geometrical spaces determine filter coefficients and pixel values determine filter system. In the two-sided filter, the output pixel values are weighted depending on the values of the neighboring pixels, wherein the weighting formula is as follows:

As shown in Eq. (2):

$$g(i,j) = \frac{\sum_{k,l} f(k,l) w(i,j,k,l)}{\sum_{k,l} w(i,j,k,l)} \quad (2)$$

where, the weight coefficients w (i, j, k, D) depend on the product of the domain kernel D (i, j, k, D) and the range kernel r(i, j, k, l), the formulas are shown in Eq. (3), Eq. (4) and Eq. (5).

$$d(i,j,k,l) = \exp\left(-\frac{(i-k)^2+(j-l)^2}{2\sigma_d^2}\right) \quad (3)$$

$$r(i,j,k,l) = \exp\left(-\frac{\|f(i,j)-f(k,l)\|^2}{2\sigma_r^2}\right) \quad (4)$$

$$w(i,j,k,l) = \exp\left(-\frac{(i-k)^2+(j-l)^2}{2\sigma_d^2} - \frac{\|f(i,j)-f(k,l)\|^2}{2\sigma_r^2}\right) \quad (5)$$

Two-sided filtering preserves image edges better by considering both spatial and value domain differences [9]. Therefore, this study will use bilateral filtering method to smooth the image to remove noise and solve the distortion problem in image segmentation.

The purpose of sharpening filter is to highlight the edge of the image and make the image clearer. By adding gradient or finite difference to the high-frequency components in the image, the edges and contours in the image are more obvious. Laplacian operator based on second-order differential is often used to achieve image sharpening.

The Laplace operator calculates pixel grayscale differences within an image neighborhood, an image enhancement technique derived from second-order differential [10]. It computes gradients in four or eight directions of the center pixel, adds these gradients to assess the relationship between the center pixel's grayscale and others, and adjusts pixel grayscale based on the gradient operation's result [11]. Its calculation formula is shown in Eq. (6).

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \quad (6)$$

C. Image Segmentation

OpenCV [12] is an open-source computer vision library, which contains rich visual processing algorithms. In terms of image segmentation, there are three classic algorithms: watershed segmentation algorithm, pyramid segmentation algorithm and mean shift segmentation algorithm [13]. Their implementation process is simple, as long as the corresponding

algorithm function can be called to complete the image segmentation according to the edge and other features.

Compared with the other two classical OpenCV algorithms, the watershed algorithm is easier to implement. However, if the input image has no obvious feature edge or is seriously affected by noise, the target area in the image will be difficult to represent, which makes the image over-segmentation phenomenon appear in the watershed algorithm based on gradient image. To compensate for this shortcoming, OpenCV provides an improved watershed algorithm that uses Markers to mark how different regional gradient-guided image segmentation is defined to effectively reduce oversegmentation [14].

OpenCV's GrabCut is a popular image segmentation algorithm. It utilizes image texture and boundary info with minimal user interaction for excellent segmentation. It's a graph-based method where each pixel is a node, and pixel dissimilarity is expressed by weighted edges. Cuts' capacity corresponds to an energy function, with min/max flow algorithms used to cut the graph. The resulting min cut corresponds to the desired boundary [15].

III. Deep Learning

As a subfield of human intelligence, deep learning uses neural networks as the main model. Convolutional neural network and capsule network are two representative network models, which are often used in image processing and image classification.

CNN is a deep feedforward network with local receptive fields, shared weights, and pooling [16]. It mainly consists of convolution, pooling, fully connected layers, and activation functions. Various combination ns of these layers create neural network models with distinct performances [17]. The network model structure is shown in Fig. 1.

A. Convolution Layer

The convolution layer, the heart of CNN, comprises several kernels with pairs of weights and biases [18]. It extracts features from input images, influenced by kernel size. Nodes in the layer receive input from the preceding network, and convolution analyzes each part deeply to yield a more abstract feature set [19].

The convolution kernel is a filter that applies to image parts based on its size, like 3×3 or 5 ×5 grids. Each channel in the convolution layer uses a distinct filter. It convolves RGB images into five feature maps, with different filter values per channel. Filter size and stride (pixels between convolutions) can vary, affecting the learned features. Images may be sampled by pixels based on layer hyperparameters and zero padding. Outputs from multiple channels can be fed into a merging layer.

B. Pool Layer

The pooling layer serves to filter and select the features extracted by the convolutional layer, effectively reducing the matrix size and subsequently diminishing the number of parameters in the fully connected layer. This is achieved by the pooling layer's ability to decrease pixel information in the input image [20]. Usually, the maximum value in each pool is used. The output result is the maximum value in each single block area. In general, the pooling layer will be connected after the convolution layer in CNN networks, because pooling can reduce the space size of volume feature data, reduce the number of parameters and calculation in the network, and suppress the occurrence of over-fitting to a certain extent.

C. Fully Connected Layer

Full connection integrates local features extracted before it into a complete graph via a weight matrix. The fully connected layer, with its multi-layered structure, acts as a “classifier” in CNN. After processing through convolution and pooling layers, extracted features contain high-level image information. Connecting the fully connected layer maps these features to the sample mark space, performs non-linear combinations, and classifies the image using the extracted features. The final classification recognition is obtained through the softmax layer.

In neural networks, receptive field maps the pixel range on the feature map from each conv layer. Traditional neural nets connect each input image pixel to a neuron, leading to a large number of weights and training difficulties. The local receptive field in CNNs depends on the conv kernel size, establishing local connections to form extracted features, reducing weights. By setting conv step size, overlapping areas are avoided, preventing weight increase.

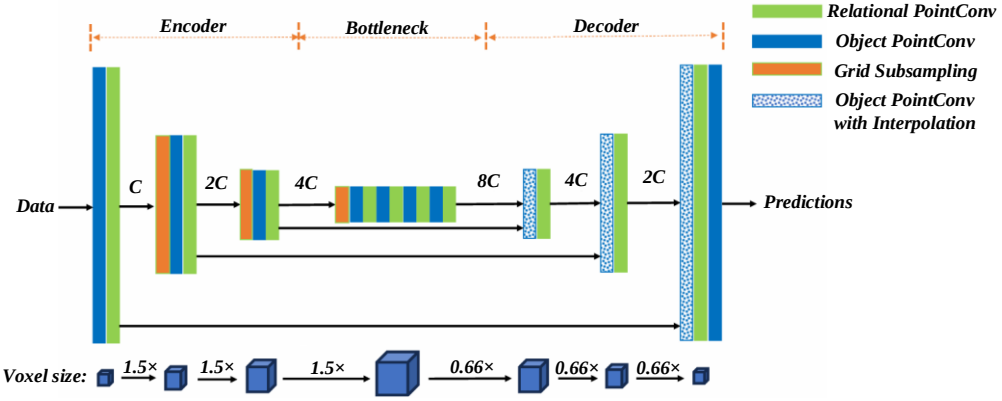

Fig. 1. Structural diagram of the network model.

The convolution kernel's weight is learned and remains constant during convolution. This ensures that the same target in different image positions exhibits similar characteristics. Weight sharing reduces the number of weights in the model. For instance, a 3×3 kernel with nine parameters convolves with different image areas to detect the same features. Different kernels correspond to unique weight parameters for detecting distinct features.

In the convolutional neural network model, the network model with different performance can be obtained by combining different number of convolutional layers and pooling layers into different network structures.

The advantages of the Inception network model are mainly reflected in the control of parameter quantity and calculation amount, while ensuring a higher classification accuracy. The Inception model replaces the full connection layer with global average pooling, reducing overfitting in the classified network. Network performance is enhanced by widening the network. Different-sized convolution kernels enrich layer information in each module. The third edition introduces convolution factorization, decomposing large kernels into smaller ones, saving parameters and reducing model size. The latest version incorporates the residual idea of ResNet for deeper networks.

The main contribution of the ResNet residual network model is the discovery of degenerative phenomena, and the invention of fast connections for degenerative phenomena, and the inclusion of congruent connections, so that gradient propagation can skip the convolution layer, even if the number of network layers reaches a thousand layers can still be trained [21]. The problem that the depth of neural network training is too large is eliminated greatly, and the problem that the learning ability of neurons decreases with the increase of the depth of the network model is solved.

D. Dynamic Routing Algorithm

The dynamic routing algorithm enables the capsule network to achieve superior recognition results. It involves capsules in lower layers predicting and learning instantiation parameters for upper layers via transformation matrices. Consistent predictions from multiple capsules activate upper-layer capsules, outputting feature vectors with expanded receptive fields. This algorithm comprises vector calculations and route selections, detailed in specific computational expressions.

The capsule layer activation output vector Vj is calculated as Eq. (7), Eq. (8) and Eq. (9).

$$S_j = \sum_i c_{ij}\, \widehat{u_{j|i}} \tag{7}$$

$$\widehat{u_{j|i}} = W_{ij} u_i \tag{8}$$

$$v_j = \frac{\|s_j\|^2}{1+\|s_j\|^2} \frac{s_j}{\|s_j\|} \tag{9}$$

Routing parameters, which are used to realize dynamic routing between capsule layers. The specific calculation is shown in Eq. (10) and Eq. (11).

$$b_{ij} \leftarrow \widehat{u_{j|i}} \cdot v_j \tag{10}$$

$$c_{ij} = \frac{\exp b_{ij}}{\sum_k \exp \widehat{b_{ik}}} \tag{11}$$

The loss function can be used to evaluate the implementation effect and performance of the model. The classical capsule network loss function adopts the interval loss function, and the specific calculation is shown in Eq. (12).

$$L_k = T_k max(0, m^+ - \| v_k \|)^2 + \lambda(1 - T_k) max(0, \| v_k \| - m^-)^2 \tag{12}$$

IV. CNN-CapsNets Classification Model

A. Model Structure

By integrating the strengths of two prominent deep learning network models, we have formulated the CNN-CapsNets model specifically for wheat disease and insect pest classification. This model is a fusion of classical convolutional neural networks, ResNet, Inception, and Capsule Networks. The CNN-CapsNets model effectively retains feature information through the utilization of the capsule layer within the Capsule Network architecture. Due to the shallow layer of the capsule network, the ability to obtain features is limited. So, in the model CNN-CapsNets we design multiple channel structures. In each channel we extract more image features through different numbers of convolution layers pooling layers and capsule combinations [22]. To mitigate under-fitting in capsule networks for large-scale images, a pooling layer is employed after convolutional feature extraction to downsize the image, thereby reducing computational parameters. Finally combined with the idea of Inception model to discard the full connection layer and implement classification in the SoftMax layer. The model structure is shown in Fig. 2.

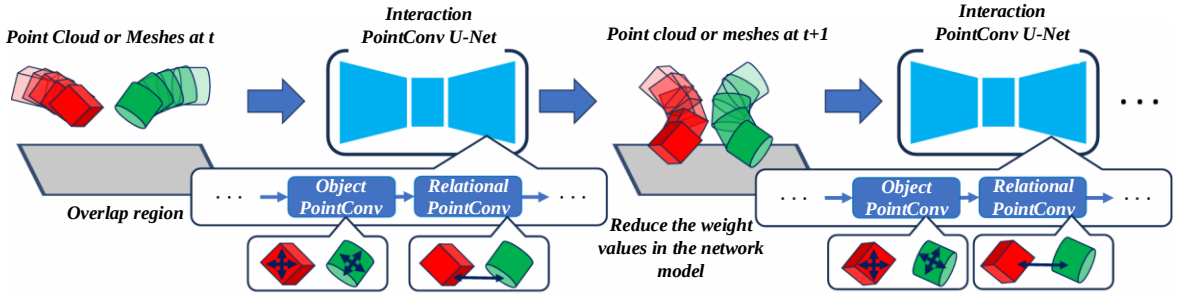

Fig. 2. Structural diagram of the CNN-Caps Nets classification model.

Fig. 2 shows that the CNN-Caps Nets model is multi-channel, dividing the Caps Net's result matrix into several parts. Parallel processing of different channels and lines enhances training efficiency.

B. Optimization Strategy

In the process of establishing the classification model of wheat diseases and insect pests, this paper takes the data set as the input sample information of the network model, and provides the following optimization strategy assumptions on training and optimizing the model. Although both convolutional neural network and capsule network have the ability to automatically extract features, the images of wheat diseases and insect pests taken in the production environment basically have complex backgrounds. If the images are segmented in advance, can the classification recognition degree of the classification model be improved? In this paper, the improved watershed algorithm and GrabCut algorithm are used to segment the image respectively, and the processed images are established respectively. The data set without image segmentation (dataset-no), the improved watershed image segmentation data set (dataset-w), and the GrabCut image segmentation data set (dataset-g). After that, the datasets were used for model training and quizzes, respectively. Finally, the performance and recognition of the classification model are taken as a reference to select the optimal image segmentation scheme.

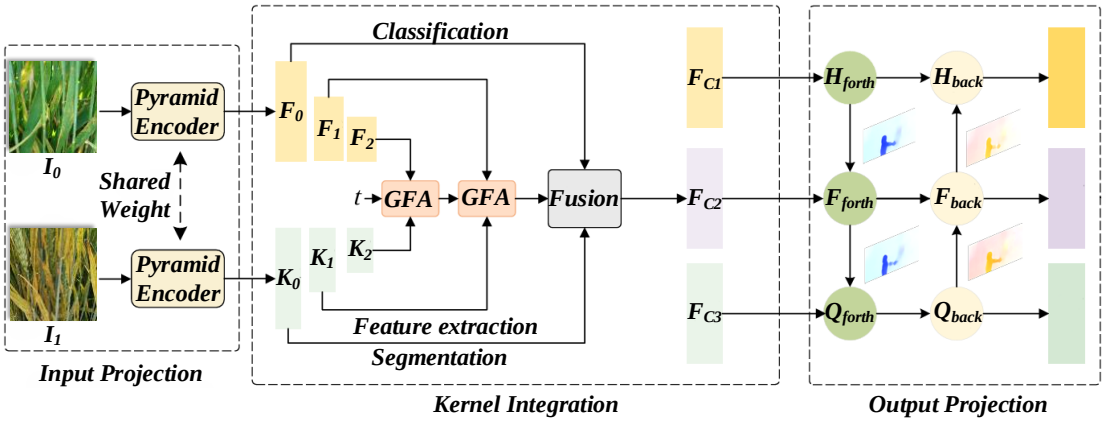

Fig. 3. Multimodal data processing and fusion process.

Fig. 3 shows multimodal data processing and fusion process. The classification model of pests and diseases in this study is designed as a multi-channel structure, which can extract features in different channels to improve the classification accuracy. Although this structure has the characteristic of parallel processing, it has a good effect on improving the efficiency of the model. However, the hardware requirements of parallel processing are also improved, and it is also necessary to consider that when the number of channels increases infinitely, the learned image features will be repeated, which will lead to redundancy in the classification model structure and affect the accuracy and efficiency of the type. Therefore, under the current hardware equipment conditions, the two-channel, four-channel, and eight-channel structure models are designed respectively. To determine the optimal number of channels, various network models with diverse architectures are trained and validated using a uniform dataset. However, it is worth noting that as the number of capsules increases, the computational load of the network model also escalates accordingly. In order to ensure a higher recognition degree and optimize the recognition efficiency of the acquired model, under the optimal number of paths, a model with 4, 8, 10, and 16 capsules in each primary capsule layer is designed. Fig. 4 shows comparison diagram of the multimodal data fusion effect.

C. Cross Validation

Cross-validation is often used as a precision test method, and its main purpose is to verify the stability of the designed network model and whether there is an over-fitting phenomenon [23]. Cross-validation is also called loop estimation. In a given training sample, most of the data is taken out for modeling, and a small part of the data is used to verify the established model. In this way, the suitable optimal network model can be found. Given the limited number of samples in the dataset, this study adopts the leave-one-out cross-validation method, where the original training set is partitioned into a new training set and a validation set in a 9:1 ratio.

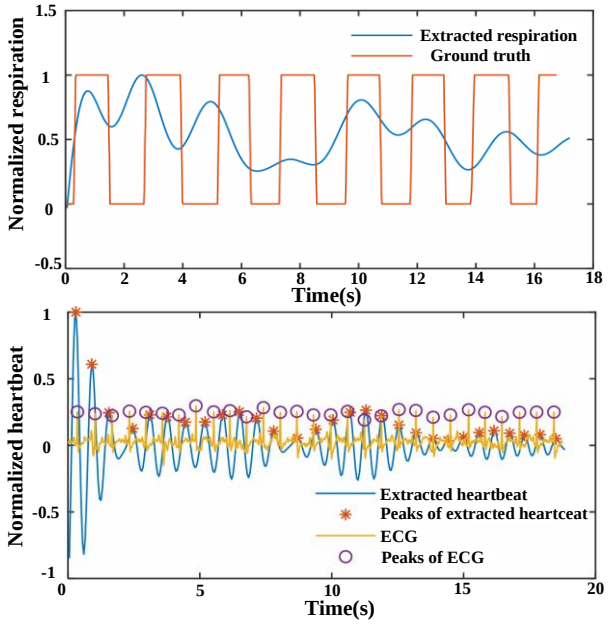

Fig. 4. Comparison diagram of the multimodal data fusion effect.

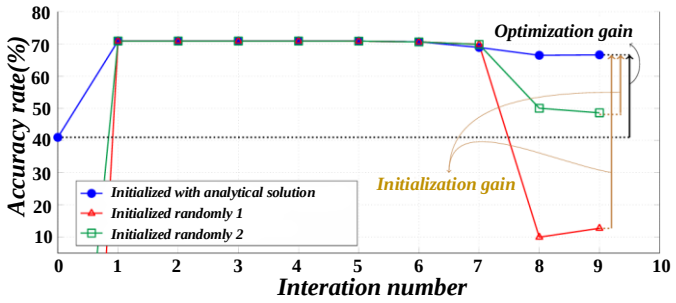

Fig. 5. Curve of image recognition accuracy of wheat disease pests over time.

Fig. 5 shows curve of image recognition accuracy of wheat disease pests over time. The training set trains the classification network, while the verification set checks for overfitting. Performance is assessed by comparing verification accuracy [24].

V. Result Analysis

A. Influence of Image Segmentation Selection on Model

In this study, the accuracy of the test set is used as the judging standard, and the reserved test set is divided into three processing methods: no image segmentation, improved watershed image segmentation, and GrabCut image segmentation to complete the test set establishment, and the data sets of different image segmentation algorithms. The model is trained, and the results are shown in the Fig. 6.

Fig. 6. results of datasets with different image segmentation algorithms. The figure indicates that the enhanced watershed and GrabCut segmentation dataset has minimal impact on enhancing the classification model's accuracy for training. Although GrabCut has a reduction in training time, when GrabCut image segmentation, the segmentation time is too long for large-size images [25]. So finally select the data set that does not segment the image in advance to train the model.

B. Influence of Channel Number on Model

The accuracy of the training set and verification set without image segmentation is used as the evaluation criterion. The training results of the network model with different channels of 2, 4, and 8 are shown in the figure, and the verification results are shown in Fig. 7. Fig. 7 shows training results of the network models for the different channels. Fig. 7 shows that as the number of model channels increases, training accuracy nears 97%, but training time also rises due to the added channels.

Fig. 8 shows distribution of identification accuracy for different wheat disease pest categories. Finally, a classification model with four channels is selected according to the verification accuracy and training time. In comparison to the other two models, the classification model with four channels exhibits the highest verification accuracy, while maintaining a relatively short training time, thus ensuring optimal training efficiency [26].

C. Effect of Capsule Quantity on Model

The training results of the network model with 4, 8, 10, 16 capsules in each primary capsule layer under the 4-channel model are shown. When using the same data set to train the network models with different capsule number structure, the training accuracy has no obvious difference, and the final training accuracy is in the range of 97 +0.45%.

Fig. 9 shows training results of the network model with different numbers of capsules in the main capsule layer. However, when comparing the verification accuracy, it can be seen that after 10 Epoch, the highest verification accuracy is the CNN-CapsNets model with 16 capsules in the primary glue layer [27, 28]. However, through continuous cycle verification, it is found that the accuracy of the four models rises gently and the accuracy begins to approach. Due to the limitation of device memory, when the number of capsules is increased again on the basis of 16, the time required to run the algorithm is too long. Therefore, based on the limitations of the current hardware equipment, considering the time and accuracy, this paper adopts a 4-channel classification model with 16 capsules.

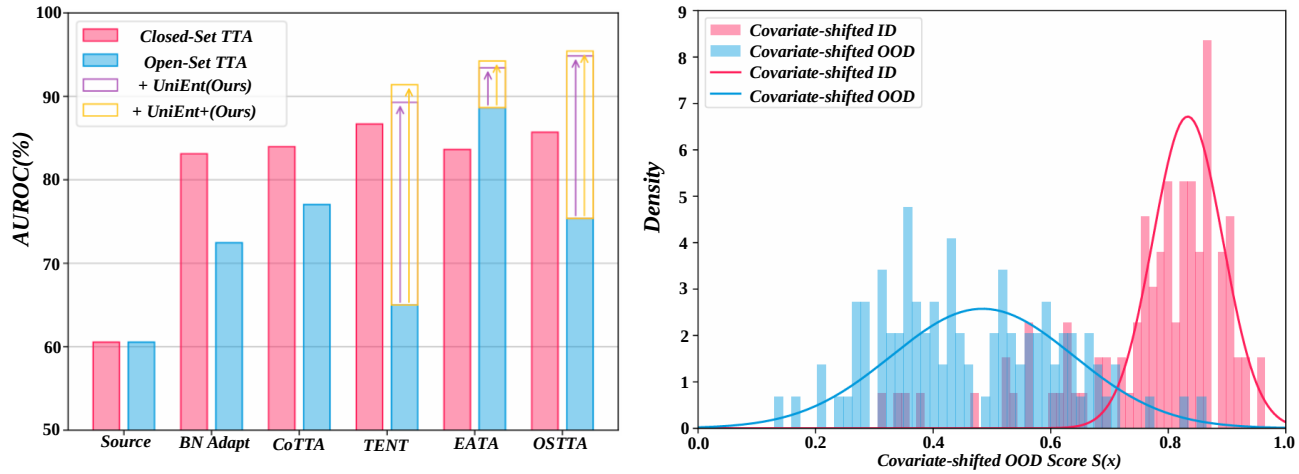

Fig. 6. Results of datasets with different image segmentation algorithms.

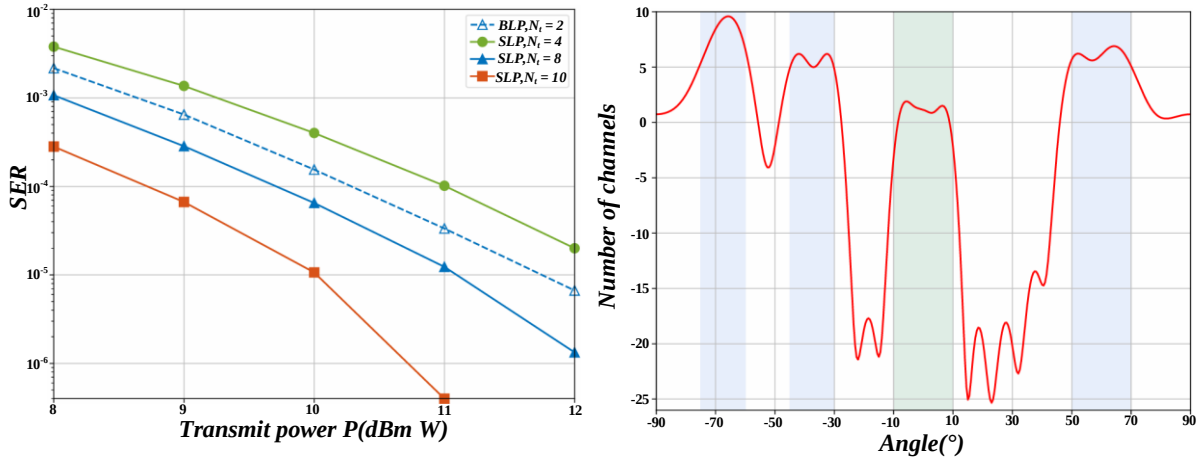

Fig. 7. Training results of the network models for the different channels.

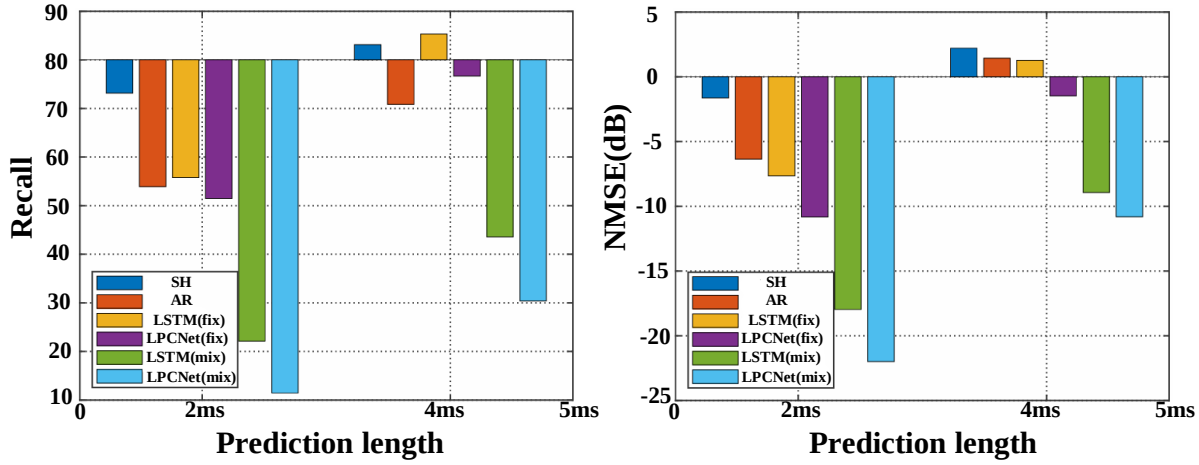

Fig. 8. Distribution of identification accuracy for different wheat disease pest categories.

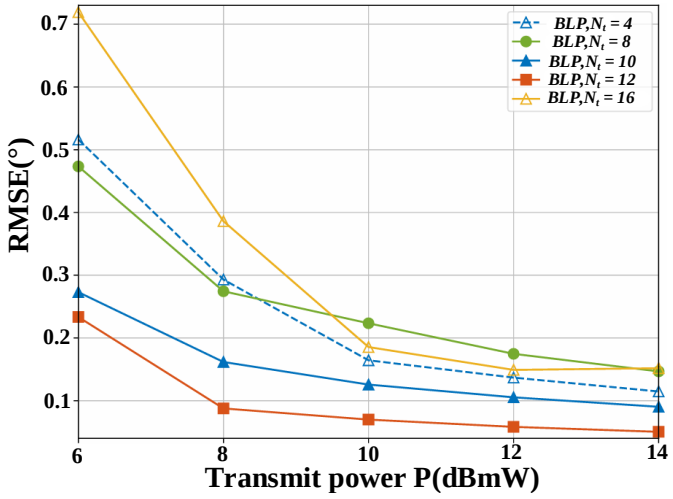

Fig. 9. Training results of the network model with different numbers of capsules in the main capsule layer.

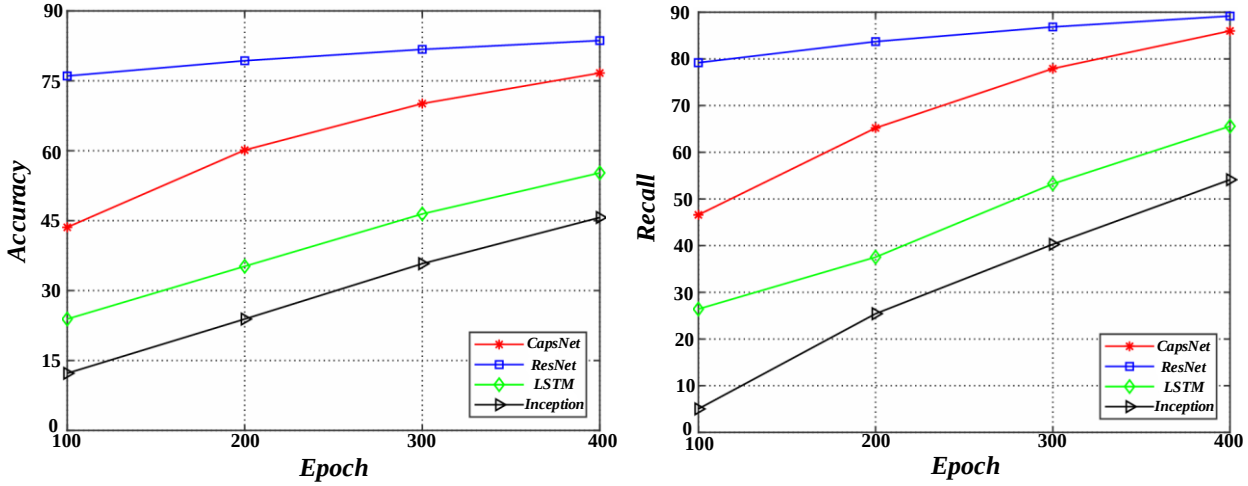

Fig. 10. Effect of multimodal data enhancement on identification performance.

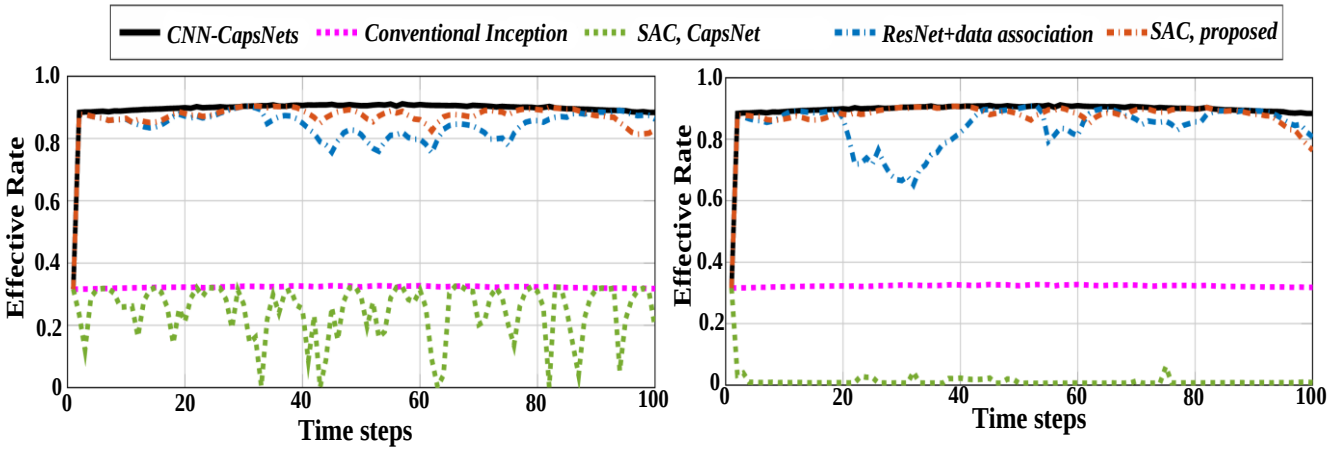

Fig. 11. Comparison of multiple model training results.

Fig. 10 shows the effect of multimodal data enhancement on identification performance. The CNN-CapsNets classification model was compared with classical models such as Inception model, ResNet model, CapsNet model, etc., and the results are shown in the Fig. 11.

As illustrated in the figure, the model employed in this study achieves a significantly superior classification accuracy compared to the Inception, ResNet, and CapsNet models. However, training time is longer than Inception and ResNet but shorter than CapsNet. This is because the CNN-CapsNets model extracts more features through multiple channels, which increases the cost of capsule layer parameter calculation and leads to increased training time. Compared with the CapsNet model, the training practice of the CNN-CapsNets model uses the convolution layer to extract image features, thereby reducing the dynamic routing computational overhead of using capsules to extract features.

The classification model is tested with a pre-dense test set, as shown in the figure. The number of training samples impacts recognition accuracy in deep learning models. More samples enhance the model's generalization and recognition accuracy [29]. Because the images in the training sample and the test sample are not pre-processed in this study, the image quality is different, which reduces the recognition rate to a certain extent. Among the four diseases, powdery mildew and leaf rust have higher recognition accuracy, not only because of the large number of samples, but also because these two diseases have prominent spot characteristics, for example, powdery mildew will appear on the surface of the plant with white powdery mildew. Mildew layer, the image features are obvious, and the network model is easier to extract features, so the classification effect is better [30].

VI. Conclusion

In this paper, four common wheat diseases and insect pests, wheat powdery mildew, wheat leaf rust, wheat scab and wheat midge, are used as research objects, combined with deep learning technology to study the classification and recognition method of pest images, and use Python, Java, and WeChat applet technology to build A wheat pest image recognition system.

To implement the classification of pests and disease images, a classification network model, termed CNN-CapsNets, is established by integrating convolutional neural networks and capsule networks. The model can extract more different features to generate feature maps through multi-channel and multi-level structural characteristics, and then save more image feature

information for classification through the high retention of capsule layer features, so the classification recognition rate is higher. Because the model can complete the internal calculation of the model in parallel in the form of multi-threads, the time required to process features is shortened, so it is faster than the classic capsule network in the same training environment.

Completed the development of the image recognition system of wheat diseases and insect pests. The migration of the pest classification model is realized by Python programming, and the API is developed to realize the call of the small program client, and the development and implementation of the identification system is completed. Finally, the recognition speed of the system is kept within 15s on average.

VII. FUNDING

This study was supported by Science and Technology Project of Henan Province (Project No. 242102111190) and Project Support for Key Scientific Research Projects of Henan Provincial University (Project No. 24B520017).

REFERENCES

[1] Lu, Y., Chen, D., Olaniyi, E., & Huang, Y. (2022). Generative adversarial networks (GANs) for image augmentation in agriculture: A systematic review. Computers and Electronics in Agriculture, 200, 107208.

[2] Stephen, A., Punitha, A., & Chandrasekar, A. (2024). Optimal deep generative adversarial network and convolutional neural network for rice leaf disease prediction. The Visual Computer, 40(2), 919-936.

[3] Kolluri, J., Dash, S. K., & Das, R. (2024). Plant Disease Identification Based on Multimodal Learning. International Journal of Intelligent Systems and Applications in Engineering, 12(15s), 634-643.

[4] Patil, R. R., & Kumar, S. (2022). Rice-fusion: A multimodality data fusion framework for rice disease diagnosis. IEEE access, 10, 5207-5222.

[5] Bhugra, S., Srivastava, S., Kaushik, V., Mukherjee, P., & Lall, B. (2024). Plant Data Generation with Generative AI: An Application to Plant Phenoty**. Applications of Generative AI, 503-535.

[6] Zhang, J., Rao, Y., Man, C., Jiang, Z., & Li, S. (2021). Identification of cucumber leaf diseases using deep learning and small sample size for agricultural Internet of Things. International Journal of Distributed Sensor Networks, 17(4), 15501477211007407.

[7] Zhang, J., Rao, Y., Man, C., Jiang, Z., & Li, S. (2021). Identification of cucumber leaf diseases using deep learning and small sample size for agricultural Internet of Things. International Journal of Distributed Sensor Networks, 17(4), 15501477211007407.

[8] Mahmoud, M. A., Guo, P., & Wang, K. (2020). Pseudoinverse learning autoencoder with DCGAN for plant diseases classification. Multimedia Tools and Applications, 79(35), 26245-26263.

[9] Li, D., Song, Z., Quan, C., Xu, X., & Liu, C. (2021). Recent advances in image fusion technology in agriculture. Computers and Electronics in Agriculture, 191, 106491.

[10] Feilong, T., Yew, H. T., Wong, F., & Porle, R. R. (2024, January). Advancements for Improved Plant Disease and Pest Identification: A Survey. In 2024 International Conference on Green Energy, Computing and Sustainable Technology (GECOST) (pp. 354-358). IEEE.

[11] Ünal, Z. (2020). Smart farming becomes even smarter with deep learning—a bibliographical analysis. IEEE access, 8, 105587-105609.

[12] Sahu, P., Chug, A., Singh, A. P., & Singh, D. (2023). Classification of crop leaf diseases using image to image translation with deep-dream. Multimedia Tools and Applications, 82(23), 35585-35619.

[13] Usha Ruby, A., George Chellin Chandran, J., Chaithanya, B. N., Swasthika Jain, T. J., & Patil, R. (2024). Wheat leaf disease classification using modified ResNet50 convolutional neural network model. Multimedia Tools and Applications, 1-19.

[14] Farooqui, N. A., Mishra, A. K., & Mehra, R. (2022). Automatic crop disease recognition by improved abnormality segmentation along with heuristic-based concatenated deep learning model. Intelligent Decision Technologies, 16(2), 407-429.

[15] Huang, X., Chen, A., Zhou, G., Zhang, X., Wang, J., Peng, N., ... & Jiang, C. (2023). Tomato leaf disease detection system based on FC-SNDPN. Multimedia tools and applications, 82(2), 2121-2144.

[16] Huang, X., Chen, A., Zhou, G., Zhang, X., Wang, J., Peng, N., ... & Jiang, C. (2023). Tomato leaf disease detection system based on FC-SNDPN. Multimedia tools and applications, 82(2), 2121-2144.

[17] Xu, K., Shu, L., Q., Song, M., Zhu, Y., Cao, W., & Ni, J. (2023). Precision weed detection in wheat fields for agriculture 4.0: A survey of enabling technologies, methods, and research challenges. Computers and Electronics in Agriculture, 212, 108106.

[18] Dai, G., Fan, J., & Dewi, C. (2023). ITF-WPI: Image and text based cross-modal feature fusion model for wolfberry pest recognition. Computers and Electronics in Agriculture, 212, 108129.

[19] Khan, A., Vibhute, A. D., Mali, S., & Patil, C. H. (2022). A systematic review on hyperspectral imaging technology with a machine and deep learning methodology for agricultural applications. Ecological Informatics, 69, 101678.

[20] Chen, Y., Huang, Y., Zhang, Z., Wang, Z., Liu, B., Liu, C., ... & Qian, W. (2023). Plant image recognition with deep learning: A review. Computers and Electronics in Agriculture, 212, 108072.

[21] Yu, H., Liu, J., Chen, C., Heidari, A. A., Zhang, Q., Chen, H., ... & Turabieh, H. (2021). Corn leaf diseases diagnosis based on K-means clustering and deep learning. IEEE Access, 9, 143824-143835.

[22] Deng, J., Zhang, X., Yang, Z., Zhou, C., Wang, R., Zhang, K., ... & Ma, Z. (2023). Pixel-level regression for UAV hyperspectral images: Deep learning-based quantitative inverse of wheat stripe rust disease index. Computers and Electronics in Agriculture, 215, 108434.

[23] Deng, J., Zhang, X., Yang, Z., Zhou, C., Wang, R., Zhang, K., ... & Ma, Z. (2023). Pixel-level regression for UAV hyperspectral images: Deep learning-based quantitative inverse of wheat stripe rust disease index. Computers and Electronics in Agriculture, 215, 108434.

[24] Patel, B., & Sharaff, A. (2023). Automatic Rice Plant's disease diagnosis using gated recurrent network. Multimedia Tools and Applications, 82(19), 28997-29016.

[25] Abdolrasol, M. G., Hussain, S. S., Ustun, T. S., Sarker, M. R., Hannan, M. A., Mohamed, R., ... & Milad, A. (2021). Artificial neural networks based optimization techniques: A review. Electronics, 10(21), 2689.

[26] Farooqui, N. A., Mishra, A. K., & Mehra, R. (2023). Concatenated deep features with modified LSTM for enhanced crop disease classification. International Journal of Intelligent Robotics and Applications, 7(3), 510-534.

[27] Ye, C. W., Yu, Z. W., Kang, R., Yousaf, K., Qi, C., Chen, K. J., & Huang, Y. P. (2020). An experimental study of stunned state detection for broiler chickens using an improved convolution neural network algorithm. Computers and electronics in agriculture, 170, 105284.

[28] Ye, C. W., Yu, Z. W., Kang, R., Yousaf, K., Qi, C., Chen, K. J., & Huang, Y. P. (2020). An experimental study of stunned state detection for broiler chickens using an improved convolution neural network algorithm. Computers and electronics in agriculture, 170, 105284.

[29] Yan, J., & Wang, X. (2022). Unsupervised and semi-supervised learning: The next frontier in machine learning for plant systems biology. The Plant Journal, 111(6), 1527-1538.

[30] Gao, J., Westergaard, J. C., Sundmark, E. H. R., Bagge, M., Liljeroth, E., & Alexandersson, E. (2021). Automatic late blight lesion recognition and severity quantification based on field imagery of diverse potato genotypes by deep learning. Knowledge-Based Systems, 214, 106723.