# COMP9444 Project Summary

# Insect Pest Specifies Identification


| | |
|---|---|
| Liran Fang | z5400642 |
| Zhikai Lai | z5412820 |
| Zihang Cheng | z5502944 |
| Yitong Li | z5504073 |
| Jinghan Wang | z5286124 |


## I. Introduction

Classification of insect pests is essential for agriculture, pest control and ecological research. Accurate and timely identification of pests is essential for effective pest management, early detection of invasive species and maintenance of crop quality. However, manual classification is often slow, error-prone, and challenging, especially in natural environments. This project enhances pest identification through the Resnet50 model MSA-Resnet model and integrated feature extraction and classification methods (including Swin-T, VAE,and SVM) to improve the efficiency and accuracy of pest management practices.

## II. Related Work

We refer to paper [1], The study evaluates the performance of state-of-the-art deep convolutional networks on the IP102 dataset, including AlexNet, GoogleNet, VGGNet-16, and ResNet-50. All networks are pre-trained on the ImageNet dataset and then fine-tuned on the IP102 dataset. Deep features are extracted from the CNNs by removing the last layer in the model architectures. Fine-tuning involves adjusting all layers of the pre-trained models using a mini-batch Stochastic Gradient Descent optimizer with specific parameters (learning rate, weight decay, and momentum). Dropout is employed to prevent overfitting. However, we found that the dataset was large, while there was a clear lack of conformity between the pictures and classes in the dataset. At the same time, these methods are inefficient, time-consuming, and perform poorly in representing advanced semantic information. So in our subsequent work we are going to take the approach of manually labeling some of the obviously problematic data content, while we believe that in the future research should be focused on coping with the high intra-class variation and class imbalance in insect pest classification. Techniques such as data augmentation, semi-supervised learning and transfer learning can be explored to improve model performance.

We also refer to the fine-grained attention mechanism (MSSN) based on the multi-scale information sharing network proposed in this paper [3], combined with the adaptive particle swarm optimization algorithm, to improve the accuracy and efficiency of agricultural pest detection. We think it is a good way to add an attention mechanism module to the model, so we define MSA-ResNet model, introduce attention mechanism module and multi-scale convolution, which may improve the accuracy of prediction.

## III. Methods

Based on the characteristics of our database, which include a large number of insect species to be identified, an imbalance in the number of images among these species, and the overall large size of the database, we utilized four common Convolutional Neural Network models to be pre-trained.

AlexNet: A simpler architecture with 8 layers which contain stacked convolutional layers followed by fully connected layers; GoogLeNet: Inception modules perform convolutions with multiple kernel sizes in parallel; Vgg16: Deep network with 16 layers, using 3x3 convolutions throughout; ResNet50:Deep residual network with 50 layers.

After varieties of experiments and tests, we got the optimal parameters for four models: learning rate set to 0.0001, momentum set to 0.9, weight_decay set to 5e-4 and different dropout for each model: 0.7 for AlexNet, 0.5 for GoogLeNet, 0.5 for Vgg16, 0.6 for ResNet. According to the results , ReNet50 and Vgg16 performed better, while the accuracy of Reset50 showed a tendency to rise and theoretically ResNet is more suitable for our project, because Vgg16 will easily overfit based on our imbalanced datasets.

Based on the previous base model, we developed an innovative model called MSA-ResNet (Multi-Scale Attention ResNet), which is a significant improvement over the traditional ResNet architecture. The core feature of this model is the combination of multi-scale feature extraction and attention mechanism, which makes it more effective to deal with complex image classification tasks. Specifically, MSA-ResNet uses convolution kernels of different sizes to capture multi-scale features in the image, so that it can focus on the local details and the overall structure of the image at the same time. We also introduce channel and spatial attention mechanisms that enable the model to automatically identify and reinforce important features while suppressing irrelevant information. It was then fine-tuned for the specific task of insect classification. To evaluate the performance of the model, we use a range of standard metrics including accuracy, precision, recall, and F1 score, among others. To further improve the model performance, we employ an optimizer to automatically tune the hyperparameters. This process ensures that the model is at its best across different training stages and thus performs well on a variety of complex image tasks.

The next model we tried was Swin-T Transformer, a model that further optimizes classification accuracy by introducing a Transformer-based architecture.

Swin Transformer performs well in high-precision and complex feature extraction tasks such as image classification, target detection, and semantic segmentation. It is well suited for applications that rely on high-performance computing resources, such as cloud computing. Due to its excellent performance, Swin Transformer is the best choice for scenarios that require the highest performance with sufficient computational resources. Future work will focus on optimizing its attention mechanism, reducing computational complexity and memory usage, and developing a lightweight version for resource-constrained environments. In addition, it is planned to extend the model to more computer vision tasks to further improve performance and generalization.

In addition, we employ a variational autoencoder (VAE) for classification. VAE models are particularly useful in scenarios where labeled data is scarce. They perform well in downscaling and feature extraction tasks and are valuable tools for the pre-processing step of classification. By generating synthetic images, VAEs can also perform data augmentation, thus increasing the diversity of the training dataset. For classification tasks, VAEs can be utilized to learn robust feature representations from large amounts of unlabeled data, which can be used more efficiently to train classifiers.VAEs are also suitable for anomaly detection by identifying anomalous patterns in the reconstruction error, and support semi-supervised learning by exploiting both labeled and unlabeled data. However, the current VAE model is oversimplified and may limit its ability to capture detailed features in images. To improve its performance, we need to increase the complexity of the model and enhance its architecture to extract more meaningful and subtle features, ultimately improving classification accuracy and overall effectiveness.

## IV. Experimental Setup

The primary data source for this project is the IP102 dataset, known as the "insect-pest-dataset," which contains a total of 75,222 images categorized into 102 different pest classes. The dataset is structured hierarchically, where each pest is assigned to a high-level category based on the crop it affects. This hierarchical classification system allows for a detailed understanding of pest-crop relationships. Key features of the dataset include its large number of samples and its representation of multiple pest species, enhancing the robustness of the study.

During data exploration, we observed significant class imbalance, a common challenge in datasets with a wide range of categories. To address this issue, we implemented targeted data augmentation techniques, which helped to balance the distribution of classes and improve model training outcomes.

For the modeling, we employed the MSA-ResNet model with an optimizer that automatically adjusts key hyperparameters such as learning rate, batch size, and dropout rate, using an SGD optimizer. To evaluate the effectiveness of our model, we utilized the accuracy metric and generated confusion matrices and heatmaps to provide detailed insights into the model's performance across different pest categories.

## V. Results

In method 1, by comparing AlexNet,Vgg16,GoogLeNet,and ResNet, we finally found that ResNett is the most suitable and its accuracy is: 71. Because of its superior migration learning performance and its ability to capture complex features, but its drawbacks are the high computational complexity and the long training and inference time.

So we tried method 2: MSA-ResNet.His data performance is shown in Fig. The final test accuracy obtained is 72.52%. Combining convolutional kernels of different scales improves feature diversity, which improves classification performance and generalization ability. Compared with ordinary ResNet combining multi-scale features and attention mechanisms, it increases the diversity and effectiveness of features and enhances the classification performance and generalization ability of the model.

Finally we tried method III: integrated feature extraction and classification method (including Swin-T, VAE,and SVM) . Its final accuracy is 74%. In this method, we combined the multi-scale feature extraction capability of Swin-T and the latent space representation capability of VAE, as well as the classification capability of SVM in high-dimensional feature space. Compared with MSA-ResNet, it can provide more comprehensive and rich feature representation, and because swin-t focuses on global, while the ResNet model is more localized. Both of them are well suited for classification, just from different perspectives.

Since all three models need to be trained, the IFCNN part has no way to classify directly since the final output is the weights, so we need to manually label the weights for each graph, and this takes too much time and effort, so we just implemented the final model but did not perform the training operation, we went through this model to learn the usage of the composite model, and the training Strategy.

## VI. Conclusions

Different models have their own advantages and disadvantages in specific tasks, and choosing the right model is crucial.Our model selection is from ResNet50, MSA-ResNet, combined with SVM for classification, while we further optimize the model performance by adjusting the learning rate, adding Dropout layer and introducing attention mechanisms. The overfitting and computational complexity problems were gradually solved, and the combination of these methods allowed us to achieve better performance on large-scale and complex datasets. In the future, we plan to further explore hybrid models and optimize the parameters to reduce overfitting and improve the generalization ability of the model.

## VII. reference

[1]. X. Wu, C. Zhan, Y. -K. Lai, M. -M. Cheng and J. Yang, "IP102: A Large-Scale Benchmark Dataset for Insect Pest Recognition," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019, pp. 8779-8788, doi: 10.1109/CVPR.2019.00899.

https://ieeexplore.ieee.org/document/8954351
https://openaccess.thecvf.com/content_CVPR_2019/papers/Wu_IP102_A_Large-Scale_Benchmark_Dataset_for_Insect_Pest_Recognition_CVPR_2019_paper.pdf

[2]. A. Setiawan, N Yudistira, and R.C. Wihandika, “Large scale pest classification using efficient Convolutional Neural Network with augmentation and regularizers”, Computers and Electronics in Agriculture, Vol. 200, Sept 2022. https://www.sciencedirect.com/science/article/pii/S0168169922005191

[3]. W. Linfeng, L. Yong, L. Jiayao, W. Yunsheng, and X. Shipu, “Based on the multi-scale information sharing network of fine-grained attention for agricultural pest detection”, PLOS ONE 18(10):e0286732. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0286732

[4]. An J, Du Y, Hong P, Zhang L, Weng X., “Insect recognition based on complementary features from multiple views”, Scientific Reports. 2023 Feb;13(1):2966. DOI: 10.1038/s41598-023-29600-1. https://europepmc.org/article/pmc/pmc9940688

[5]. S. Kar, J. Nagasubramanian, D. Elango, M. E. Carroll, C. A. Abel, A. Nair, D.S. Mueller, M. E. O’Neal, A. K. Singh, S. Sarkar, B. Ganapathysubramanian, A. Singh, “Self-supervised learning improves classification of agriculturally important insect pests in plants”, The Plant Phenome Journal, 6, e20079. https://acsess.onlinelibrary.wiley.com/doi/full/10.1002/ppj2.20079