ICPR 2024 Competition on Beyond Visible Spectrum: AI for Agriculture ⋆

Liangxiu Han[1], Wenjiang Huang[2], Xin Zhang[1], Yingying Dong[2], Tam Sobeih[1], and Yufan Lin[3]

[1] Department of Computing, and Mathematics, Manchester Metropolitan University, Manchester M15 6BH, UK
[2] Key laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China
[3] Institute of Data Science, National Cheng Kung University, Taiwan 70101, ROC
l.han@mmu.ac.uk

Abstract. The ICPR 2024 Competition on "Beyond Visible Spectrum: AI for Agriculture" presents an exciting opportunity for researchers and practitioners to advance computer vision techniques in agricultural crop disease monitoring. By leveraging the vast multi/hyper spectral remote sensing datasets, participants are encouraged to develop innovative machines/deep learning algorithms. The primary objective is to improve the accuracy and efficiency of crop disease identification, thereby fostering sustainable agricultural practices and facilitating global knowledge generation for enhanced food security. This challenge presents participants with the opportunity to make significant contributions to the advancement of precision farming and crop monitoring techniques. By improving the machines/deep learning models, the competition aims to drive forward the capabilities of agricultural technology, making it more efficient, accurate, and sustainable. A total of 190 people from more than 10 countries in Europe, Asia and North America entered the competition by the deadline date, with a total of 126 submissions of results. The rank1 team obtained an accuracy of 73.3% in crop disease classification task with hyperspectral dataset. In the self-supervised learning task, the rank1 team obtained 79.2% accuracy. We report the leader-boards of the competition in this report and present three innovative approaches of the participants. The ICPR 2024 Competition on "Beyond Visible Spectrum" is not just a competition; it is a call to action for the computer research community to bring about transformative changes in how agricultural data is analyze and utilized, paving the way for smarter, more sustainable farming practices that can benefit communities worldwide.

Keywords: Remote sensing · Crop disease · Hyperspectral · Deep Learning · Self supervised learning.

⋆ Supported by Manchester Metropolitan University and Chinese Academy of Sciences. https://han-research.gitlab.io/Agvision/

1 Introduction

The ICPR 2024 Competition on "Beyond Visible Spectrum: AI for Agriculture" aims to encourage researchers and practitioners to push the boundaries of computer vision techniques in the agricultural sector, particularly in the area of crop disease monitoring[5]. With the advent of advanced multi/hyperspectral remote sensing image datasets, the competition seeks to foster the development of deep learning algorithms that can significantly enhance the accuracy and efficiency of identifying crop diseases. These advancements are expected to contribute to more sustainable agricultural practices and bolster global food security[6].

Crop diseases and pests are responsible for 20 to 40 percent of the world's crop yield loss each year. According to the Food and Agriculture Organization, the annual cost of crop diseases and pests to the global economy is estimated at $220 billion, posing a critical challenge to food security and the agricultural economy. Early, accurate diagnosis and quantification of crop diseases are crucial for ensuring the stability of food supplies and the livelihoods of millions of people involved in agriculture. However, traditional monitoring methods are labor-intensive, time-consuming, and often impractical on a large scale.

The advent of remote sensing technology has provided new ways of monitoring agriculture. With continuous global monitoring, it is possible to monitor vast agricultural areas in real-time, making it possible to identify and address threats before they cause significant damage. With the recent advancement in digital imaging, a variety of image-based methods that perform automated image acquisition and analysis are developed for plant management and have shown great potential for automated crop disease diagnosis based on images from RGB (red, green, and blue), thermal, multi-spectral, and hyperspectral sensors. Particularly, hyperspectral imagery (HSI) carries more narrow spectral bands over a contiguous spectral range, which provides detailed spectral–spatial information of the disease infestation, offering the potential to provide better diagnostic accuracy[8].

Deep learning methods have shown promising results in various tasks related to remote sensing imagery analysis. Identifying relevant spectral features and extracting meaningful representations from remote sensing data are crucial for effective deep learning-based analysis. However, since the number of spectral bands in remote sensing images is significantly higher than that of traditional RGB images, and the relationship between them is complex, feature selection and extraction are extremely challenging and require the development of new deep learning algorithms [9].

The primary objective of the competition is to drive innovation in precision farming and crop monitoring technologies, while also fostering global knowledge generation to enhance food security. By improving deep learning models for remote sensing data, the competition aims to advance agricultural technology, making it more accurate, efficient, and sustainable. Additionally, the competition serves as a call to action for the research community to explore new methods for analyzing and utilizing agricultural data, with the ultimate goal of enabling smarter and more sustainable farming practices that can benefit communities worldwide.

The competition comprises two tasks: **Task 1** focuses on hyperspectral imagery, while **Task 2** involves multispectral imagery. Task 1 aims to detect the Fusarium Head Blight (FHB) by leveraging joint spectral and spatial information extracted from provided hyperspectral imagery. This task involves classifying the hyperspectral data into two categories: Mild FHB and Serious FHB. In Task 2, participants are required to utilize self-supervised learning techniques to identify various types of crop diseases from remote sensing imagery. This task addresses the issue of limited labeled data in remote agricultural regions and encourages the development of models capable of learning meaningful representations without extensive annotated datasets[10, 7].

The paper is structured as follows. In Section 2, we present the details of the tasks and datasets used in the challenge. Section 3 discusses participants' results and evaluation procedures, whereas Section 4 concludes the current status of the challenge.

2 The challenge setup and dataset

The competition comprises two tasks: **Task 1**: Automated Crop Disease Diagnosis from Hyperspectral Imagery and **Task 2**: Boosting automatic crop diseases classification using Sentinel satellite data and self-supervised learning. In this section we provide separate descriptions of the task setup and the dataset for each task.

Task 1: Automated Crop Disease Diagnosis from Hyperspectral Imagery

This taks focuses on hyperspectral imagery to detect Fusarium Head Blight (FHB) by leveraging joint spectral and spatial information extracted from the provided hyperspectral imagery. This task involves classifying the hyperspectral data into two categories: Mild FHB and Serious FHB.

The datasets for Task 1 consist of a collection of hyperspectral datasets specifically curated for the detection of wheat stripe rust disease. The hyperspectral imagery retrieval is carried out in two times, on 3 May and 8 May 2019. On 3 May 2019 wheat in the experimental area is in the pre-grouting stage, which is a critical period for the spread of downy mildew infestation. On 8 May 2019 wheat in the experimental area is in the middle of the grouting stage, which is the period when the symptoms of downy mildew are more obvious.

A DJI M600 Pro UAV system (SZ DJI Technology Co Ltd., Gungdong, China) with a snapshot hyperspectral sensor is used for data acquisition. The model of the hyperspectral sensor is S185, which can obtain reflect radiation from the visible to near-infrared band between 450-950nm wavelength. The spectral resolution is 4nm. The raw data is recorded as a 1000*1000 panchromatic image and a 50*50 hyperspectral image with 125 bands. Due to the interference caused

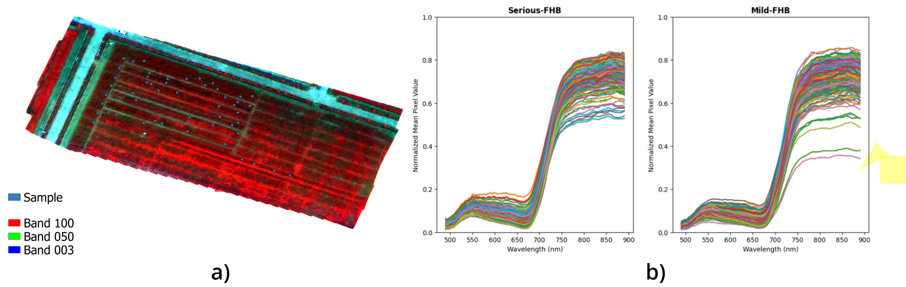

Fig. 1. a) the acquired image in false colour and sample position, b) the spectrum profiles of the Mild FHB and Serious FHB

by noise during the post-processing of UAV hyperspectral data, which significantly affects the data at both ends of the spectrum, the first 10 bands and the last 14 bands are excluded from the study, resulting in a dataset of 100 usable bands. In this work, all the images are obtained at an altitude of 60 meters, with a spatial resolution close to 4cm per pixel. Figure 1 a) shows the the acquired image in false colour and sample position.

In task 1 the image covers areas of FHB plots. We partition the image with a size of 32*32*100 and labell them into two categories according to its location: Mild FHB and Serious FHB. Figure 1 b) shows the spectrum profiles of the Mild FHB and Serious FHB.

Task 2: Boosting automatic crop diseases classification using Sentinel satellite data and self-supervised learning

Participants are challenged to employ self-supervised learning techniques to identify various types of crop diseases from remote sensing imagery. This task addresses the issue of limited labelled data in remote agricultural regions and encourages the development of models capable of learning meaningful representations without extensive annotated datasets. For this challenge, substantial remote sensing datasets covering arable land areas, including multi-spectral and synthetic aperture radar data (Sentinel 1/2), are provided. This dataset contains 500,000 images for participants to train their self-supervised models. Then, a crop disease classification task and datasets will be established covering Wheat stripe rust, blight, aphids, and health. Participants will use this data to fine-tune their self-supervised models. The competition will determine the performance of a self-supervised model by evaluating how much its performance improves over a model that does not use the self-supervised training dataset.

The datasets for Task 2 consist of a collection of remote sensing data on cropland areas. The datasets include unlabelled and labelled patches (Sentinel-2 surface reflectance multi-spectral) covering the wheat major production zones in China. At each location, 4 images are captured from four seasons to cover

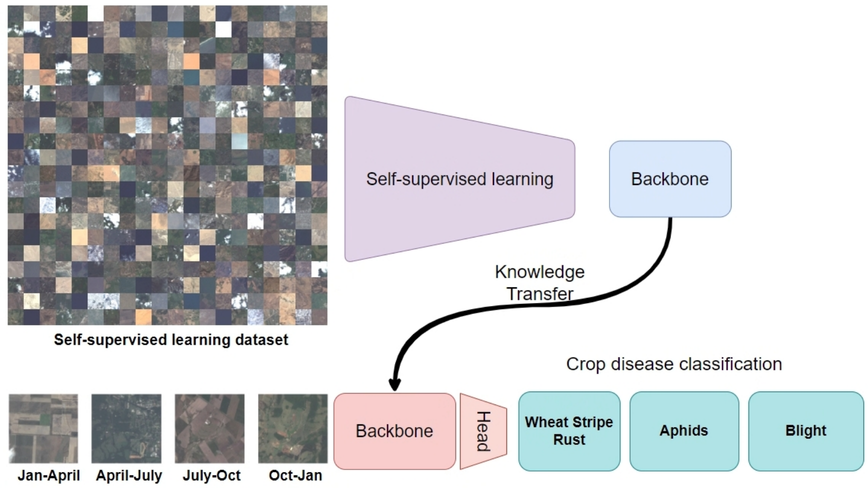

Fig. 2. The dataset used in Task 2: Boosting automatic crop diseases classification using Sentinel satellite data and self-supervised learning

the different growing seasons of the crop (Jan-April, April-July, July-Oct, and Oct-Jan). Each patch has a size of 280*280, covering 2800m x 2800m.

In task 2, we provide both unlabelled and labelled data patches. The unlabelled patches are remote sensing images of cropland areas provided to the participants for self-supervised pre-training. The labelled patches contain three types of crop disease and health (Wheat stripe rust, blight, and aphids). These data are used to perform supervised training and evaluation.

2.1 Data Setup

For this competition, training data and test data are provided. The training data is used to allow participants to train and test their models. The test data is only provided as data, with the truth labels stored on the server to automatically generate evaluation results. In Task 1, 80% of the data is training data, and 20% of the data is test data. In Task 2, the training data consists of self-supervised training data and supervised training data. 500,000 unlabelled self-supervised training data patches and 100,000 labelled data patches are provided. 80% of the labelled data is used for training, and 20% is used for testing.

3 Competition Results and Analysis

3.1 Quality Metric

Categorization Accuracy is defined as the ratio of correctly categorized items to the total number of items evaluated. It is a measure of how accurately a classification system or model assigns items to their correct categories based on a predefined standard.

$$\text{Categorization Accuracy} = \frac{\text{Number of Correctly Categorized}}{\text{Total Number}} \times 100\% \quad (1)$$

3.2 Summary of Competitor Algorithms

A total of 190 people with 28 teams participated in this competition and developed different methods for this competition. In this section, we review and discuss some methods proposed by the top competitors to solve the classification task in task 1 and tasks 2. We'll present three innovative methods and summarise methods with existing models.

Task 1 solution: Bands selection with machine learning method

The team NCKU_ACVLAB introduces a strategy to effectively utilize the spectral information of image-level Hyper-spectral Imagines (HSIs) using the simple endmember extraction strategy with top-K bands selection. Afterward, the simple classifier is employed to detect the given HSI is suffered from mild-FHB or serious-FHB. The overall pipeline of the method is shown in Figure 3.

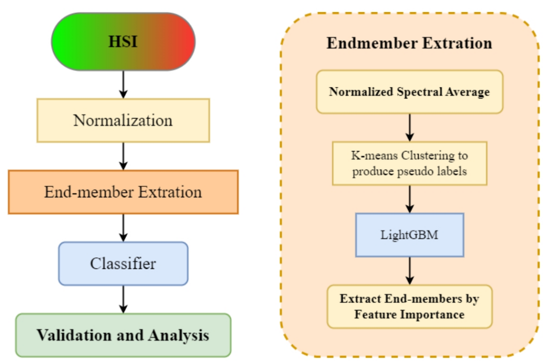

Fig. 3. The overall pipeline of the proposed method.

Top-k bands selection for Endmember Extraction

Given the unique characteristics of HSIs, where different materials exhibit distinct reflectance values at the same bands, the team proposes an efficient approach for FHB detection that bypasses the need for complex HSI unmixing techniques traditionally used in endmember extraction. Before performing endmember extraction, normalization and spectral averaging techniques are applied to the hyperspectral imagery (HSI) data. These steps are crucial for reducing data complexity and eliminating noise and redundancy, thereby improving the quality and reliability of the subsequent analysis. Because of the FHB detection can be regarded as the type of coarse level HSI classification, there is no need for keeping fine-grained feature well. Therefore, normalization strategy is beneficial to the better performance and robustness for FHB detection. However, the traditional metrics for diseased crops recognition, such as NDVI and NGRDI, are insufficient to handle more diverse scenes and the complexity of hyperspectral imagery (HSI). To reduce the complexity of HSI and streamline the recognition procedure, The team introduces a top-K band selection method for key bands extraction. The team begins by utilizing K-means clustering to generate pseudo-labels, providing valuable guidance. The normalization and spectral averaging techniques enable K-means clustering to achieve excellent clustering results.

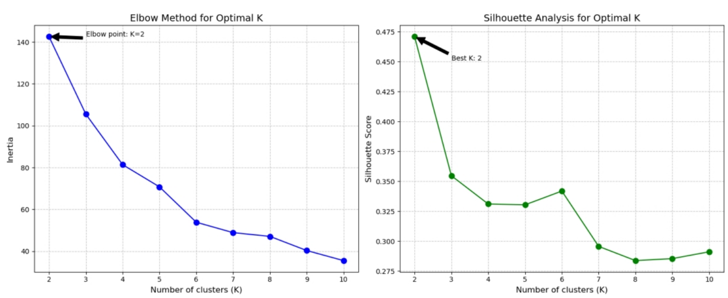

Fig. 4. Elbow method(left) and Silhouette Analysis(right). Both methods show that when k=2 is the best selection.

To determine the optimal number of clusters (K), The team employs two complementary methods: the Elbow Method and Silhouette Analysis[1], as illustrated in Figure 4. The Elbow Method, shown in the left graph, plots the inertia (within cluster sum of squares) against the number of clusters. The optimal K is identified at the "elbow" point where the rate of decrease in inertia begins to level off, occurring at K=2 in their analysis. Correspondingly, the Silhouette Analysis, depicted in the right graph, measures how similar an object is to its own cluster compared to other clusters. The highest Silhouette score indicates the best clustering, which also occurs at K=2 in our dataset. The consistency between these two methods reinforces our confidence in selecting K=2 as the optimal value. Then, the team utilizes K-means clustering to generate pseudo-labels, providing guidance for our analysis. Due to normalization and spectral averaging techniques, K-means clustering achieved excellent clustering results. These pseudo labels are subsequently used for feature importance mining. Based on this analysis, they select the top-K important bands to serve as endmembers for mild or

K-means -> chia HS thành k=2 cm, theo chin lc hin ti
Dùng 2 cm này ánh giá quan trng ca tng bands
(band nào giúp phân bit tt 2 lp) ~ 30 bands

serious FHB recognition. Unlike traditional endmember extraction methods in HSI analysis, which primarily rely on optimization-based approaches with dense mathematical computations, our method is simpler. Its simplicity stems from the direct extraction of discriminative features, streamlining the process while maintaining effectiveness.

Simple Classifier for Fusarium Head Blight Detection

Because the robust and compact features are yielded, the classifier can be arbitrary. The team uses LightGBM as primary classifier because LightGBM excels in handling high-dimensional data with efficiency and robustness. It is especially beneficial in detecting FHB infections using HSI, which is usually a high-dimensional cube data, helping to mitigate the need for chemical treatments for treating wheat in the future.

Task 1 solution: DC2Net

The team Jason Z88 introduces a novel deep learning-based model, DC2Net, designed for the image classification using hyperspectral imaging (HSI)[4]. The proposed model, DC2Net, integrates deformable convolution and dilated convolution techniques to effectively extract spatial and spectral features from hyperspectral images, which traditional models struggle to achieve. DC2Net demonstrates superior performance compared to state-of-the-art (SOTA) models.

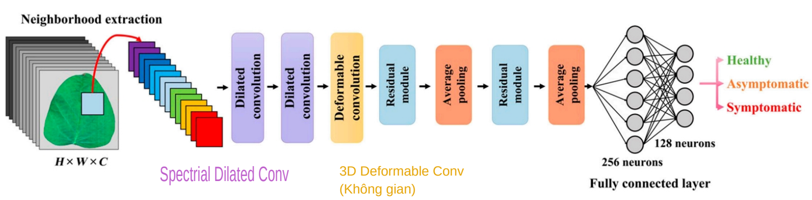

Fig. 5. The architecture of the DC2Net model

DC2Net is composed of several key components, each designed to address the challenges of hyperspectral image classification, including Spectral Dilated Convolution Modules, 3D Deformable Convolution Module and Residual Modules. The architecutre of the DC2Net is shown in Figure 5.

The Spectral Dilated Convolution Module is used to extract features in the spectral dimension. Dilated convolutions introduce gaps within the convolutional kernels, allowing the model to increase the receptive field without adding extra parameters. This helps in capturing detailed spectral information, particularly in identifying troughs or peaks in the spectral data that are indicative of crop disease.

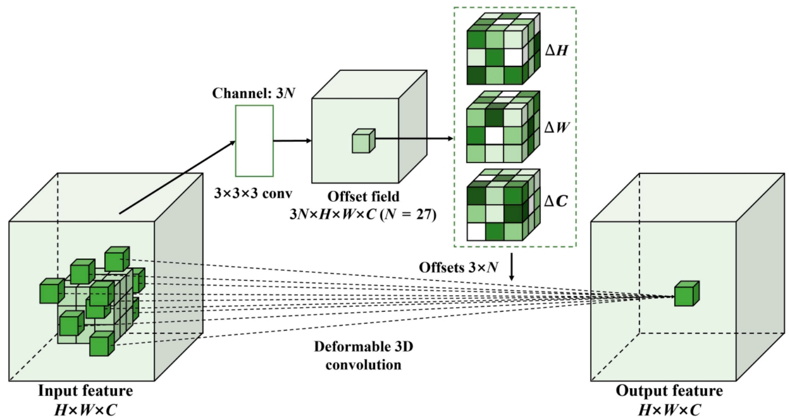

Fig. 6. The flowchart of the 3D deformable convolution module

Figure 6 shows the flowchart of the 3D deformable convolution module The 3D Deformable Convolution Module is designed to adaptively capture spatial features, particularly the irregular and scattered rust spots on crop leaves. Unlike standard convolutional kernels, deformable convolutional kernels can adjust their receptive fields dynamically by learning offsets during training. This allows the model to focus on relevant spatial regions, even if they are not aligned with the fixed grid of standard convolutions.

The Residual Modules is used to prevent the vanishing gradient problem and improve training stability.

Task 2 solution: Spectral Awareness Self-Supervised Learning

The team JustLetMeSignUp introduces a general Spectral Awareness Self-Supervised Learning (SASSL) for feature extraction of multi/hyper spectral remote sensing data. the SASSL is a contrastive learning method to learn the representation of remote sensing data. The idea of contrastive learning is to learn representations that bring similar data points (positive pairs) closer while pushing randomly selected points further away or to maximize the contrastive-based mutual information lower bound between different views (negative pairs). The pretext task of SASSL is a classification problem that uses contrastive loss to measure how well the model can classify the representation among a set of unrelated negative and positive samples. In this work, the positive samples are generated by discerning the representation of augmented views of the same data. The negative pairs assume that different images in a batch during model training represent different categories. The flowchart of the work is shown in Figure 7. There are two main parts to SASSL:

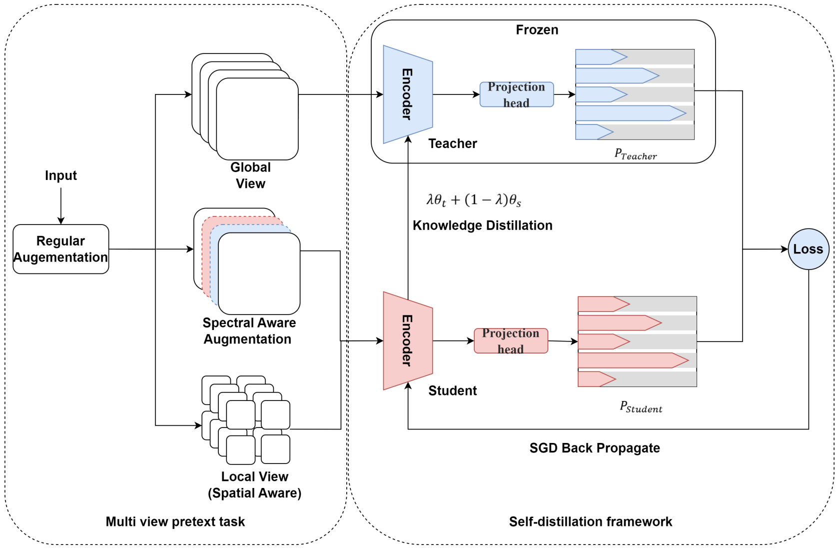

Fig. 7. Spectral Awareness Self-Supervised Learning architecture

1. A novel multi-view pretext task that generates positive pairs for SASSL by generating different views of remote sensing data from both the spectral and spatial perspectives. This is a composition of multiple data augmentation operations, including spectral aware augmentation, regular augmentation, and local and global augmentation.
2. A self-distillation framework that uses two networks, a student network and a teacher network, to learn the representation from multiviews of the data. The student network is trained to match the output of a given teacher network.

Multiview Pretext Task

In SASSL, the positive pairs are generated by applying data augmentation to create noise versions of the original samples. Appropriate data augmentation is essential for learning good, generalizable embedding features. It introduces changes to the original images without modifying the semantic meaning, thus encouraging the model to learn the essential features. The author of [2] demonstrated that the composition of multiple data augmentation operations is crucial in defining the contrastive prediction tasks that yield effective representations. In this work, a joint spatial–spectral-aware multiview pretext task is proposed to generate positive pairs of data for SASSL. This task consists of a composite of multiple data augmentation operations, including (1) regular augmentation, (2) local and global spatial augmentation, and (3) spectral aware augmentation. Regular augmentation includes common data transformations, such as random rotation and zooming, Gaussian blur, and random noise. Local and global augmentation is used to generate views of different spatial areas. With an input data X of size 120^2, the output of this augmentation is a set containing global views and several local views of smaller resolutions. The team assume that the original data contains the global context. The small crops are called local views that use an image size of 36^2. This covers less than 50% of the global view but they assume that it contains the local context. Then, the two views are fed into the self-distillation framework. All local views are passed through the student network while only the global view is passed through the teacher network. This encourages the student network to interpolate context from a small cropped image and the teacher network to interpolate context from a bigger image. Spectral-aware augmentation is a data transformation that is perform in parallel with local and global augmentation. The traditional color-based augmentation method is a set of random transformations on random channels, including variations between channels, which inevitably change the spectral order of their relative positions. In this work, the spectral-aware augmentation process drops the random channels (30–50%) and replaces them with a value of zero. This guarantees that the relationship and relative position of the different channels do not change. This view is passed through the student encoder. This encourages the student network to learn the full spectral context from the teacher network.

Application using existing methods In addition to proposing innovative approaches, several existing vision deep learning models are applied to this competition, including:

- **Yolov8x**: YOLO (You Only Look Once) is a popular real-time object detection system. YOLOv8x is a variant of the YOLO series, where "x" indicates an extra-large model designed for higher accuracy. YOLO models are known for their speed and efficiency in detecting and classifying multiple objects within an image. YOLOv8x uses a convolutional neural network (CNN) that processes the entire image in one go, making predictions for bounding boxes and class probabilities simultaneously.
- **EfficientNet-B5**: EfficientNet is a family of convolutional neural networks designed for high accuracy with fewer parameters and lower computational cost. EfficientNet-B5 is one of the variants in this family, where "B5" refers to a specific scaling of depth, width, and resolution. EfficientNet uses a compound scaling method that uniformly scales all dimensions (depth, width, and resolution) of the network to achieve better performance.
- **VGG16**: VGG16 is a well-known deep convolutional neural network. It is characterized by its simplicity, consisting of 16 layers (13 convolutional layers and 3 fully connected layers). VGG16 is known for using small 3x3 convolutional filters stacked on top of each other, which allows the model to capture more complex features. Despite its relatively large number of parameters, VGG16 is a popular choice for image classification tasks due to its strong performance.

3.3 Summary of Competition Results and Analysis

In this section, we summarise the competition results of the participants, we first report the leaderboard of the competition and then we analyse the innovative methods uploaded by the competitors.

The submitted entries are ranked based on the metric as outlined in Section 3.1. The private leaderboard is calculated with the test data. Table 1 reports the leaderboard of task 1: Automated Crop Disease Diagnosis from Hyperspectral Imagery.

Rank	TeamName	Score	SubmissionCount	Method
1	NCKU_ACVLAB	73.3%	8	Machine learning LightGBM
2	Jason Z88	71.7%	4	Deap Learning D2Net [4]
3	VietNamese	71.7%	8	Machine learning Random Forest
4	maimimib	71.7%	12	Deap Learning SE-Inception-Resnet
5	DieNeuronaleCrew	70.0%	4	Deap Learning CNN
6	Roberto Chang S	68.3%	8	Machine learning
7	Khao Cha	68.3%	2	Deap Learning CNN
8	early_bird	68.3%	4	Deap Learning CNN
9	AgMod	66.7%	3	Deap Learning EfficientNet
10	Roshan Rateria	66.7%	2	Deap Learning CNN
11	Enchanters	66.7%	2	Deap Learning CNN
12	JustLetMeSignUp	65.0%	1	Deap Learning Transformer
13	Brian Junior	63.3%	1	Machine learning Random Forest
14	Niraj	63.3%	1	Machine learning Random Forest

Table 1. The leaderboard for task 1: Automated Crop Disease Diagnosis from Hyperspectral Imagery

A total of 108 participants participated in this competition, with 56 submissions. The participants employed two main approaches for this task: traditional machine learning techniques and deep learning methods. The traditional machine learning approach typically involves band selection and feature extraction, where specific characteristics or patterns within the hyperspectral data are identified and used to train a model. This process often requires manual intervention and domain expertise to select relevant features that are most indicative of the desired output classes. On the other hand, deep learning methods, such as convolutional neural networks (CNNs)[4] and Vision Transformer (ViT) [3], automatically learn complex feature representations from raw data without the need for manual feature extraction. These methods leverage large datasets and computational power to identify subtle patterns and correlations in the hyperspectral images, often leading to higher classification accuracy and more robust models. Both approaches have their advantages and trade-offs, depending on the specific requirements and constraints of the classification task at hand. The Team NCKU_ACVLAB gets the best results in the private test data, they get an accuracy of 73.3%. In the NCKU_ACVLAB's solution, they believe that the number of selected bands is important, it may lead the different extracted feature and prediction. Two complementary methods, the Elbow Method and Silhouette

Analysis are used for band selection. Their experiments reveal that selecting 30 bands result in perfect accuracy on the validation set. The Team Jason Z88 introduces a noval deep learning model Deformable and Dilated Convolutional Neural Networks (DC2Net). The model integrates both deformable and dilated convolutional modules to enhance feature extraction from hyperspectral images. The deformable convolution module adapts to various shapes and sizes of disease spots by adjusting the sampling locations, making it more effective in capturing irregular patterns in the spatial dimension. The dilated convolution module expands the receptive field without increasing the number of parameters, allowing it to capture more relevant spectral features, particularly in complex data with multiple wavelengths. They have obtained 73.1% accuracy in the test data.

Rank	TeamName	Score	SubmissionCount	Method
1	Vietnamese	79.2%	10	Pretrained Model
2	JustLetMeSignUp	70.8%	2	Self Supervised Learning
3	ElyyLong Truong	62.5%	21	Pretrained Model
4	In Processing	62.5%	13	Pretrained Model
5	Nguyen Hoang Anh	58.3%	1	Self Supervised Learning
6	Long Truong	58.3%	2	Self Supervised Learning
7	ElonMusk113	58.3%	1	Pretrained Model
8	Thanhnh1	57.5%	3	Pretrained Model
9	Pradeep Dalall	54.2%	6	Pretrained Model
10	oiKamP	52.4%	1	Pretrained Model
11	HoangThanhNguyen	52.4%	2	Pretrained Model
12	david32141	51.7%	2	Self Supervised Learning
13	aloalo12345	51.7%	2	Self Supervised Learning
14	AgriRG	51.7%	4	Self Supervised Learning

Table 2. The leaderboard for task 2: Boosting automatic crop diseases classification using Sentinel satellite data and self-supervised learning

Table 2 reports the leaderboard of task 2: Boosting automatic crop diseases classification using Sentinel satellite data and self-supervised learning. A total of 82 participants participated in this competition, with 70 submissions. 14 teams achieved more than 50% accuracy on the test dataset. There are two main approaches commonly used: pretrained models and self-supervised learning. Pretrained models are deep learning models that have already been trained on a large dataset, typically on a standard benchmark dataset like ImageNet. These models have learned a wide range of features and patterns from the data, which allows them to perform well on similar tasks without needing to be trained from scratch. Self-supervised learning, on the other hand, is a technique where the model learns to understand the underlying structure of the data without relying on labeled examples. Instead, it uses the data itself to generate labels, creating tasks such as predicting the position of patches within an image or reconstructing parts of an image that have been deliberately masked. Both pretrained models and self-supervised learning offer unique advantages for this task.

Pretrained models provide a head start by utilizing existing knowledge, making them highly effective when data and computational resources are limited. Self-supervised learning, meanwhile, offers a powerful way to leverage vast amounts of unlabeled data, potentially leading to models that generalize better. The team JustLetMeSignUp proposes a novel SASSL framework. In SASSL, a novel multiview pretext task is proposed to generate representations from unlabeled data. In their experiments, the result demonstrate that the representations generated from self-supervised learning improve the performance of downstream tasks. After pre-training with self-supervised learning, the deep learning model can converge faster and better in supervised training with a limited dataset.

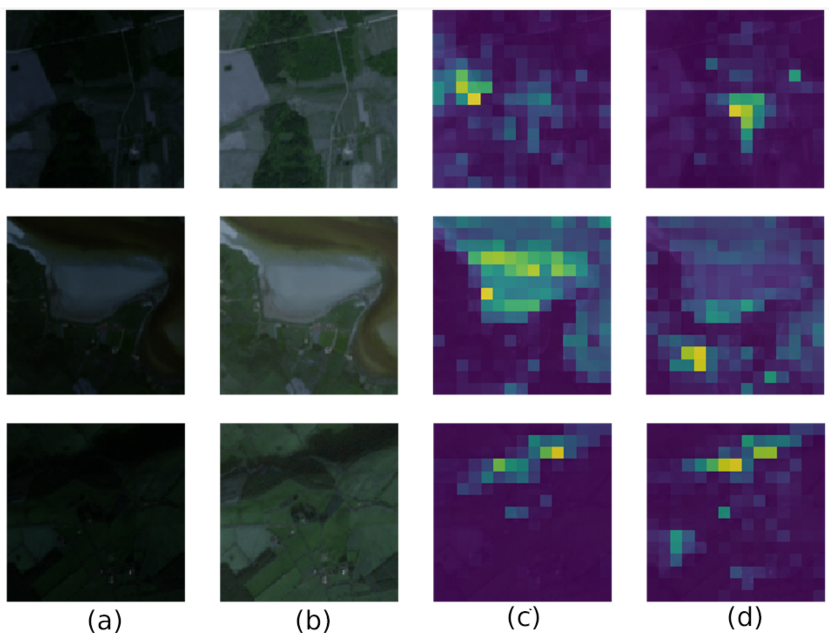

Fig. 8. Visualization of the attention maps from the last layer in the encoder. a) and b) shows the RGB vision of the data with their enhancements. c) and d) visualize the different attention maps of the last layer in the encoder after SASSL.

Figure 8 visualizes the attention maps for the different heads of the last layer of the encoder after SASSL. The (a) column is the original data displayed by red, green, and blue channels. In the (b) column, they adjust the brightness of the image for better display. The (c) and (d) columns visualize the different attention maps of the last layer in the encoder after SASSL. The results show that the attention map can attend to different semantic regions of an image, which demonstrates that the representations obtained by SSL reflect the semantic information of the data.

As demonstrated in this competition, the methods participants used presents significant opportunities for enhancing crop monitoring and disease management.

The models explored in this research, which utilize hyperspectral imagery and Sentinel multi spectral data, have shown great promise in improving the accuracy and efficiency of crop disease detection. By leveraging detailed spectral and spatial feature of the data, the models can detect and classify crop diseases with high precision. This capability is crucial for early intervention, allowing farmers to apply targeted treatments that can prevent widespread crop loss. The models trained on massive datasets can be made to have better scalability and adaptability. They can be tailored to different agricultural scales—from small farms to large industrial operations—and adapted to various crops and regions. This adaptability ensures that a wide range of agricultural contexts can benefit from the advancements in AI-driven crop monitoring.

While the current models have demonstrated considerable accuracy and potential, future developments can further enhance their capabilities, particularly through the effective use of large unsupervised datasets. In the future, we plan to release larger datasets that encompass a wider range of regions, crop types, and seasonal variations to enhance the diversity and inclusiveness of unsupervised datasets. This will provide a robust foundation for model training. Additionally, we will release our own pre-trained models, allowing users to quickly utilize them for efficient training and deployment on their specific tasks. Furthermore, we plan to explore distributed computing methods and model optimization strategies, such as pruning and quantization, to reduce computational overhead. These strategies will be crucial in adapting our models for deployment in resource-constrained environments typical of large-scale agricultural operations.

4 Conclusion

The ICPR 2024 Competition on “Beyond Visible Spectrum: AI for Agriculture” provides a challenging yet rewarding platform for participants to push the boundaries of computer vision techniques in the agricultural application with remote sensing imagery. Throughout the competition, participants leveraged multi/hyper spectral remote sensing image datasets to develop and test advanced machine and deep learning algorithms aimed at improving the accuracy and reliability of crop disease monitoring.

The outcomes of the competition highlights significant advancements in the field, particularly in the areas of algorithmic innovation and dataset utilization. There are a total of 190 people entered the competition by the deadline date, with a total of 126 submissions of results. The participants from diverse backgrounds contributing a wide variety of innovative models and their insights into multi/hyper spectral data. These diversity of approaches not only demonstrates the potential of AI in transforming agricultural practices but also underscores the importance of continued research and development in this area.

Looking ahead, the methodologies and findings from this competition are expected to have a lasting impact on both the academic community and the agricultural industry. The knowledge gained here sets the stage for further advancements in precision agriculture, with the potential to significantly enhance global food security. In conclusion, the ICPR 2024 competition has not only met its objectives of driving innovation in agricultural AI but has also laid a strong foundation for future work in this critical field.

Acknowledgements This work is supported by Manchester Metropolitan University and and Chinese Academy of Sciences.

References

1. Ashari, I.F., Nugroho, E.D., Baraku, R., Yanda, I.N., Liwardana, R., et al.: Analysis of elbow, silhouette, davies-bouldin, calinski-harabasz, and rand-index evaluation on k-means algorithm for classifying flood-affected areas in jakarta. Journal of Applied Informatics and Computing **7**(1), 95–103 (2023)
2. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A Simple Framework for Contrastive Learning of Visual Representations. arXiv:2002.05709 [cs, stat] (Jun 2020), http://arxiv.org/abs/2002.05709, arXiv: 2002.05709
3. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. arXiv:2010.11929 [cs] (Oct 2020), http://arxiv.org/abs/2010.11929, arXiv: 2010.11929
4. Feng, J., Zhang, S., Zhai, Z., Yu, H., Xu, H.: Dc2net: An asian soybean rust detection model based on hyperspectral imaging and deep learning. Plant Phenomics **6**, 0163 (2024)
5. Shi, Y., Han, L., González-Moreno, P., Dancey, D., Huang, W., Zhang, Z., Liu, Y., Huang, M., Miao, H., Dai, M.: A fast fourier convolutional deep neural network for accurate and explainable discrimination of wheat yellow rust and nitrogen deficiency from sentinel-2 time series data. Frontiers in Plant Science **14**, 1250844 (2023)
6. Shi, Y., Han, L., Han, L., Chang, S., Hu, T., Dancey, D.: A latent encoder coupled generative adversarial network (le-gan) for efficient hyperspectral image super-resolution. IEEE Transactions on Geoscience and Remote Sensing **60**, 1–19 (2022)
7. Zhang, X., Han, L.: A generic self-supervised learning (ssl) framework for representation learning from spectral–spatial features of unlabeled remote sensing imagery. Remote Sensing **15**(21), 5238 (2023)
8. Zhang, X., Han, L., Dong, Y., Shi, Y., Huang, W., Han, L., González-Moreno, P., Ma, H., Ye, H., Sobeih, T.: A deep learning-based approach for automated yellow rust disease detection from high-resolution hyperspectral uav images. Remote Sensing **11**(13), 1554 (2019)
9. Zhang, X., Han, L., Han, L., Zhu, L.: How well do deep learning-based methods for land cover classification and object detection perform on high resolution remote sensing imagery? Remote Sensing **12**(3), 417 (2020)
10. Zhang, X., Han, L., Sobeih, T., Lappin, L., Lee, M.A., Howard, A., Kisdi, A.: The self-supervised spectral–spatial vision transformer network for accurate prediction of wheat nitrogen status from uav imagery. Remote Sensing **14**(6), 1400 (2022)