AgriEngineering

MDPI

Article

Segmentation of Wheat Rust Disease Using Co-Salient Feature Extraction

Hirra Anwar [1,*], Haseeb Muhammad [2], Muhammad Mohsin Ghaffar [3], Muhammad Ali Afridi [2], Muhammad Jawad Khan [1,4], Christian Weis [3], Norbert Wehn [3] and Faisal Shafait [2,5,*]

[1] School of Mechanical and Manufacturing Engineering, National University of Sciences and Technology (NUST), Islamabad 44000, Pakistan; jawad.khan@smme.nust.edu.pk
[2] School of Electrical Engineering and Computer Science (SEECS), National University of Sciences and Technology (NUST), Islamabad 44000, Pakistan; huhammad.bscs22seecs@seecs.edu.pk (H.M.); mafridi.bee20seecs@seecs.edu.pk (M.A.A.)
[3] Microelectronic Systems Design Research Group, University of Kaiserslautern-Landau, 67663 Kaiserslautern, Germany; mohsin.ghaffar@rptu.de (M.M.G.); christian.weis@rptu.de (C.W.); norbert.wehn@rptu.de (N.W.)
[4] Department of Electrical Engineering, Prince Sattam Bin Abdul Aziz University, Al-Kharj 16245, Saudi Arabia
[5] Deep Learning Laboratory, National Center of Artificial Intelligence, Islamabad 44000, Pakistan
* Correspondence: hirraanwar.pg@smme.edu.pk (H.A.); faisal.shafait@seecs.edu.pk (F.S.)

Abstract: Wheat Stripe Rust Disease (WRD) poses a significant threat to wheat crops, causing substantial yield losses and can result in total crop damage if not detected early. The localization of WRD-infected areas is a labor-intensive and time-consuming task due to the intricate and varied nature of the disease spread, especially for large plantations. Hence, segmentation of wheat crops is vital for early identification of the WRD-affected area, which allows for the implementation of targeted intervention measures. The state-of-the-art segmentation technique for WRD using the real-world semantic segmentation NWRD dataset is based on a UNet model with the Adaptive Patching with Feedback (APF) technique. However, this implementation is complex and requires significant resources and time for training due to the processing of each patch of the dataset. Our work in this paper improves the state-of-the-art by using a two-stage model: a Vision Transformer (ViT) classifier to distinguish between the rust and non-rust patches and a less complex co-salient object detection (Co-SOD) model for segmentation of the classified images. The Co-SOD model uses multiple rust patches to extract contextual features from a group of images. By analyzing multiple patches of wheat rust disease simultaneously, we can segment disease regions more accurately. Our results show that the proposed approach achieves a higher F1 score (0.638), precision (0.621), and recall (0.675) for the rust class with 5× less training time as compared to the previous works.

Keywords: wheat rust disease; semantic segmentation using co-salient object detection; wheat rust disease segmentation; disease localization; deep learning

check for updates

Academic Editor: Lilong Chai

Received: 15 November 2024
Revised: 13 January 2025
Accepted: 17 January 2025
Published: 23 January 2025

Citation: Anwar, H.; Muhammad, H.; Ghaffar, M.M.; Afridi, M.A.; Khan, M.J.; Weis, C.; Wehn, N.; Shafait, F. Segmentation of Wheat Rust Disease Using Co-Salient Feature Extraction. *AgriEngineering* **2025**, *7*, 23. https://doi.org/10.3390/agriengineering7020023

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Crops are the foundation of global food security, serving as the primary source of nutrition for humans worldwide. One of the major challenges in the field of agriculture is the early detection of crop diseases. Early disease detection is essential not only for saving the crop from disease spread but also for preventing economic losses. Wheat is a staple crop and the impact of the disease on wheat crops has far-reaching consequences for food security, the livelihoods of farmers, and crop productivity. Several fungal diseases

affect the wheat crop, with rust being the most common [1,2]. Three types of rust diseases exist, including leaf rust, stem rust, and stripe rust [3]. Wheat Stripe Rust Disease (WRD) is the most infectious one, which produces yellow-colored stripes on the wheat leaves. In severe outbreaks, this disease has the potential to cause yield losses of up to 100% [4]. Annually, rust diseases result in an average yield reduction of approximately 15.04 million tons, equivalent to loss of around USD 2.9 billion [5]. Hence, early detection of the WRD disease is crucial, such that infected areas are isolated and pesticides and fungicides are precisely used in specific locations to limit the spread, aiding in effective disease control before it becomes widespread.

The early stages of WRD display distinct spatial patterns and are confined to specific areas and can be visually identified by an expert. This approach requires the experts to manually visit and inspect large wheat plantations, making the process labor-intensive and time-consuming. Most previous research on WRD has focused on classification-based solutions [6–10], which can only be used for WRD classification in an image without localization. In contrast, Deep Neural Network (DNN)-based semantic segmentation can be used to estimate disease spread to precise locations. This estimation allows for an accurate mapping of the affected areas and continuous monitoring of further spread. Meanwhile, agricultural experts can quarantine the infected plants to contain the disease, thereby protecting healthy crops and ensuring better overall crop management.

Although the semantic segmentation models are effective in the identification of the localized areas, these models require a comprehensive real-world dataset to achieve reliable results. Due to this reason, in our previous work, we presented a semantic segmentation dataset for WRD called the NUST Wheat Rust Disease (NWRD) dataset [11]. Furthermore, we used the encoder–decoder-based UNet model alongside the Adaptive Patching with Feedback (APF) technique for segmentation of the WRD. Due to the high resolution of the images, these images were divided into smaller patches. Although small patches decrease input size, they increase the number of operations, resulting in significant computational demand and training time. This extensive training process has a significant environmental impact, as it consumes a large amount of power, contributing to carbon emissions [12].

Early-stage disease detection presents significant challenges for semantic segmentation models, especially when the diseased region is minimal and blends into its surrounding healthy areas. These models often struggle to deliver reliable outcomes due to the subtle and nuanced field conditions typical of early disease stages, which differ significantly from the more distinct patterns observed in advanced stages. Additionally, the presence of clutter, background noise, and varying lighting conditions further complicates accurate segmentation, as these factors introduce noise and ambiguity into the input data, reducing model effectiveness. Context plays a critical role in disease detection, as understanding the relationship between the diseased region and its surrounding environment can significantly enhance segmentation accuracy. Enhanced feature extraction methods that incorporate contextual information can help distinguish subtle disease patterns from background noise, improving model performance in complex scenarios. Our work leverages the strength of Co-SOD models, which are focused on the context of a group of images to find common and salient objects from the images. Our proposed method ensures that the model focuses on the diseased regions while ignoring irrelevant information, making it highly effective for tasks involving subtle and early-stage disease detection.

We propose a two-stage pipeline for WRD, presented as the Rust Detection Module (RDM). First, we use a Vision Transformer (ViT) to classify rust and non-rust patches in the augmented-NWRD dataset. Second, we use the co-salient object detection (Co-SOD) model for the segmentation of the rust disease. These models excel at identifying common and salient objects from a group of images, enabling more accurate segmentation [13,14] with

less complexity as compared to conventional segmentation models. For high-resolution images, using both classification and segmentation models reduces the training time as compared to using only the segmentation model. This combined approach helps the model to converge faster and reduces the computations for segmentation by filtering the non-rust patches in the classification phase. For this reason, in this work, we consider training time a metric for estimating model complexity. Overall, this paper presents the following research contributions:

1. Since the first phase of the RDM requires binary classification, it is crucial to employ a classifier with a high F1 score. Due to this reason, we evaluate multiple classifiers such as CatBoost, EfficientNet, XceptionNet, ResNet50, VGG16, and ViT on the augmented-NWRD dataset. The ViT classifier outperforms other models by achieving an F1 score of 0.83.
2. For the segmentation phase, we conducted an evaluation of different co-saliency-based segmentation models such as DCFM, GCoNet, and GCoNet+ by using F1 score, Intersection over Union (IoU), and training time evaluation metrics. On the basis of the achieved results, we present a two-stage model called RDM, which consists of a ViT classifier and DCFM model for segmentation.
3. We also show that the proposed RDM pipeline achieves a higher F1 score and Intersection over Union (IoU) with 5× less training time as compared to the previous works on the augmented-NWRD dataset.

The rest of the research article has the following structure. The related work concerning recent advancements in the fields of segmentation of WRD in particular and segmentation of crop diseases in general is presented in Section 2. Section 3 details the materials and methods adopted in this research for segmentation of WRD. It provides in-depth details of the RDM and techniques used to conduct the study and the experiments conducted. Section 4 has complete details of the quantitative and qualitative results, along with the comparisons. Section 5 presents the discussion on the results, followed by the conclusion.

2. Related Work

In recent years, researchers have been trying to solve precision agriculture problems by using computer vision techniques. The use of deep learning (DL) in the agricultural domain offers several advantages, including real-time monitoring, reduced labor costs, and the ability to process large volumes of data quickly. However, the DL models require a substantial volume of training data under various temporal and visual conditions to train more robust and resilient models. Due to this, using DL for crop management for sustainable farming has been an active area of research lately. In this section, we provide a comprehensive review of previous works focused on the classification and segmentation of WRD, as well as on the classification and segmentation of diseases in various other crops.

2.1. Classification and Segmentation of WRD Using Deep Learning Models

Li et al. used the GhostNet-V2 model for detection of wheat rust severity [15]. Using a channel shuffling layer and a Fused-MBConv alongside the GhostNet architecture allowed for an accurate severity assessment by precisely estimating the percentage of lesion areas on wheat leaves. Niu et al. [16] used an ML-based K-means clustering algorithm to segment common wheat plant diseases. The RGB images were converted to the Lab color space. Clustering was then carried out by calculating the absolute difference between each pixel and the cluster center within the Lab color space.

Liang et al. proposed an enhanced U-Net framework for segmenting wheat powdery mildew disease. The framework integrates a pyramid pooling module in the down-sampling stage to capture multi-scale features and combine them into a global representa-

tion. Experimental results showed that the enhanced U-Net achieved an mIoU of 91.4% on the wheat powdery mildew spore image dataset. Bukhari et al. presented a benchmark segmentation dataset for individual wheat ears and designed a Fully Convolutional Network (FCN) model to effectively segment wheat ears in field environments [17]. Furthermore, an advanced segmentation algorithm for Fusarium Head Blight (FHB) was presented, leveraging a pulse-coupled neural network (PCNN) optimized using K-means clustering and an improved IABC algorithm to fine-tune PCNN parameters for accurately identifying diseased spots.

Zhang et al. presented a UNet++ model to segment whole wheat ears and associated wheat scap diseased areas [18]. A custom CNN network and connected domain method were used to count all wheat ears and diseased wheat ears, respectively. Li et al. used the Octave-UNet model for the segmentation of WRD, where pore-level segmentation was carried out and achieved an mIoU of 83.44% [19].

Fang et al. present a CNN model for the classification of WRD for the LWCDC dataset. They proposed an Inception-ResNet-CE (IRCE) model with multiple attention mechanisms, which provides better results for disease detection as compared to other CNN models. With a comparison of seven CNNs, their model achieved 98.76% accuracy [8]. Li et al. presented a custom Vision Transformer (ViT)-based model called PMVT with an enhanced ViT encoder and convolution kernel. This approach helped to extract long-distance pixel-to-pixel dependencies. The proposed model achieved higher accuracy and used fewer parameters as compared to the MobileNet and SqueezeNet models [20].

Singh et al. proposed LeafyGAN, which is an augmentation technique based on the pix2pix-GAN for leaf disease extraction and generation of high quality images [21]. The authors used the Mobile-Vit model to classify the images as diseased or healthy. Zhang et al. proposed a visual Large Language Model (LLM) for detection of WRD [22]. They utilized the SAM model for segmenting the disease regions of the wheat leaves, forwarded the diseased leaves to the LLM, and generated prompts as to whether the leaves are diseased or healthy. Furthermore, the authors also presented a mobile application to model this whole process in real time. Deng at. al used a transformer based SegFormer_ViT_B5 model for segmentation of WRD. The authors presented a dataset for autumn wheat leaves [23]. They also compare their results with PSPNet, DeepLabV3, and OCRNet, and out of these, SegFormer yielded the best results for WRD segmentation.

2.2. Segmentation of Crop Diseases Using Deep Learning Models

Li et al. conducted regional segmentation to detect diseases in grape leaves using data collected via agricultural drones [24]. For segmentation, they employed a multi-fusion U-Net model combined with VGG-19. Image preprocessing involved grayscale conversion and grid point positioning techniques. Their feature extraction process emphasized the identification of leaf contours, disease contours, and the categorization of disease distribution. Kerkech et al. proposed a CNN model for detecting mildew disease in vineyards using UAV imagery [25]. Their approach integrated visible and infrared data to classify pixels into categories such as healthy, symptomatic, ground, or shadow. The method demonstrated high accuracy in disease detection at both grapevine and leaf levels. Wang et al. [26] also used FCN for maize leaf disease segmentation and achieved segmentation accuracy of 96.26%.

Wang et al. proposed a custom FCN for plant disease and pest segmentation [27]. Their method utilized convolution layers to extract multi-layer feature information from maize leaf lesion images, followed by a de-convolution layer to restore the original image size and resolution. The approach achieved an accuracy rate of 95.87%. Lin et al. used a U-Net model to segment cucumber leaves affected by powdery mildew [28]. They integrate

a batch normalization layer after each convolution layer to reduce sensitivity to weight initialization. The method achieved an average pixel accuracy of 96.08%.

Stewart et al. used a Mask R-CNN model to detect Northern Leaf Blight (NLB) lesions in a UAV-based maize image dataset [29]. The model accurately detected and segmented individual lesions, achieving an average accuracy of 96%. The IoU was 73% with a threshold of 0.50. Wang et al. utilized Faster R-CNN and Mask R-CNN for tomato disease identification and segmentation [30]. Mask R-CNN was employed for segmentation, while Faster R-CNN classified the disease class. The model demonstrated the ability to distinguish 11 tomato disease classes with high accuracy and speed. Using Mask R-CNN, the detection rate for all tomato disease classes reached 99.64%, effectively capturing the location and morphology of the affected areas. Automatic vine disease detection was proposed by [25] by using visible and infrared images obtained from two different sensors. Initially, the pixel-wise superposition of both RGB and infrared images was carried out, followed by segmentation using the deep learning SegNet model [31].

3. Materials and Methods

The proposed WRD disease detection pipeline comprises several processing stages such as patch generation, data augmentation, RDM, and result concatenation, as shown in Figure 1. In the subsequent sections, we will discuss their methodologies and underlying algorithms in detail.

3.1. Patch Generation

The images of the NWRD dataset are comprised of high-resolution (4000 $\times$ 6016 pixels) images. The processing of these high-resolution images requires a lot of memory and computing resources. Furthermore, if these large-size images are fed directly to a model during the training process, the training time will be much longer. Hence, in the first step of the pipeline, we split these images into small patches of 224 $\times$ 224 pixels. This approach significantly reduces the memory and computational footprint. Furthermore, patch-based training can improve model performance by increasing the effective receptive field, allowing the model to capture more detailed features [32].

3.2. Data Augmentation

The NWRD dataset, being a real-world dataset, exhibits class imbalance for the diseased rust class, as presented in Table 1. The dataset is dominated by the non-rust patches, which can lead to biased learning of the model. Consequently, this impacts the F1 score of classification and segmentation models, particularly in scenarios where the non-rust class is heavily outnumbered by the rust class. Our analysis of labeled images of the NWRD dataset revealed a stark disparity in class distribution, with the rust class constituting approximately a mere 10% of the entire dataset in contrast to the non-rust class.

To address this class imbalance, we employed a two-step approach. Firstly, we applied data augmentation techniques to increase the number of rust patches. We only used the training set of the NWRD dataset for augmentations; the original dataset was divided into train, val, and test sets in a ratio of 0.85:05:10, respectively. We performed six different augmentations, including vertical, horizontal flip, vertical, horizontal shear, contrast adjustment, and rotation, as shown in Figure 2. We then filtered out any augmented rust patches where the rust area was no longer visible due to the transformations. We filtered out rust patches by using a threshold of 150 rust pixels for a patch to be considered a rust patch. Overall, our training set had an original set of rust patches plus an augmented set of rust patches with six transformations. As a second step, to match the number of rust and non-rust patches in the training dataset, we reduced the non-rust patches, decreas-

ing their quantity to match the number of rust patches in the training set. This ensured that the training dataset had an equal number of rust and non-rust patches for the best experimental results.

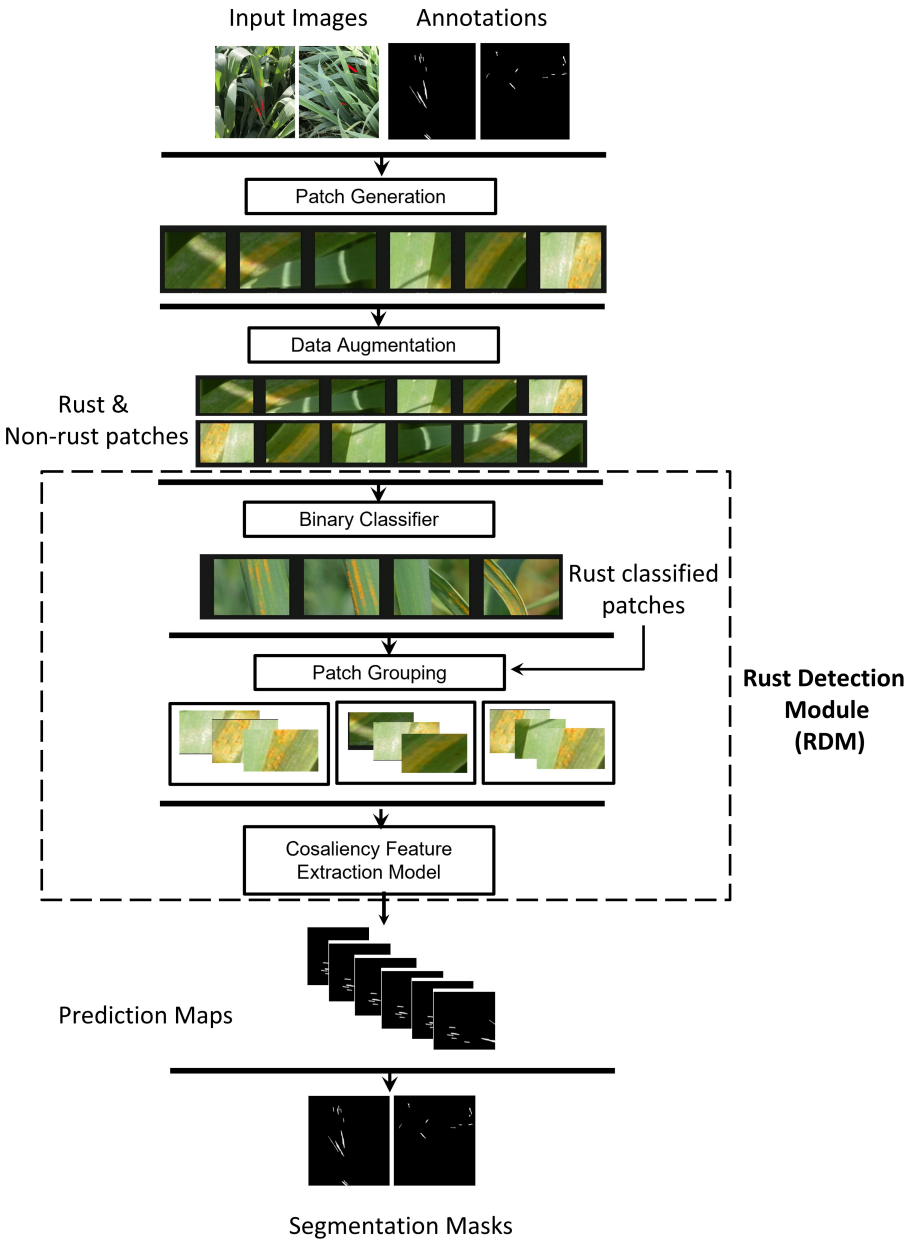


Figure 1. Our proposed WRD detection pipeline. Input images and their annotations are initially pre-processed, divided into patches, and augmented. These patches are then forwarded to the RDM for WRD localization processing, which is carried out through a classifier and a Co-SOD model. Classification separates input patches into rust and non-rust patches. Rust patches are then grouped and processed by the co-salient feature extractor to produce the final segmentation maps, which are concatenated to form the segmentation masks.

It is important to mention here that we preemptively partitioned the dataset before augmentation to address potential data leakage, i.e., unintentional exposure of information from the validation and test set during the data augmentation process, which can lead to the overly optimistic performance estimates of the deep learning models. The number of patches for the original data and augmented data are shown in Table 1.

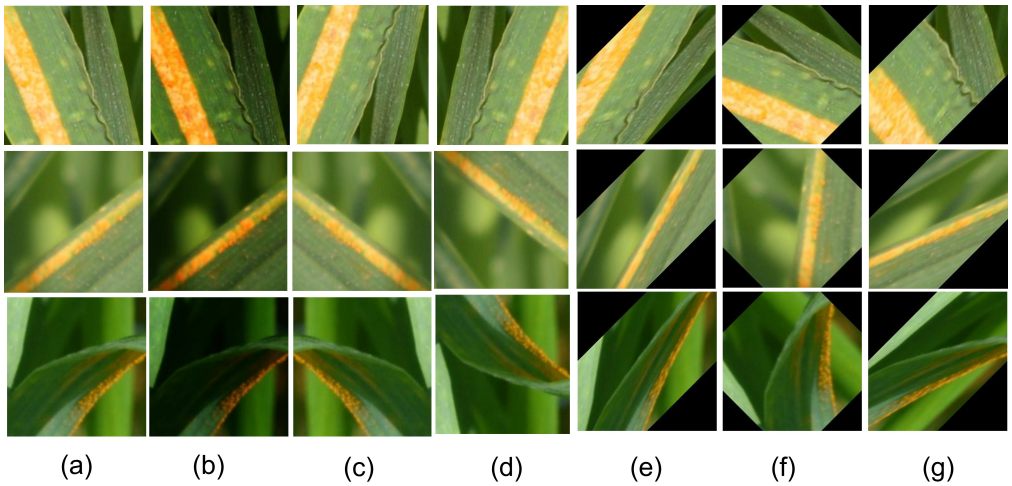


Figure 2. Sample images of the data augmentation process. (**a**) The original rust patch and (**b**–**g**) augmented data after transformations.

Table 1. Breakdown of the patches generated for the NWRD dataset and the augmented-NWRD dataset during the process of training, validation, and testing of the classifier and Co-SOD model.

	NWRD		**Augmented-NWRD**	
	Rust	**Non-Rust**	**Rust**	**Non-Rust**
Training	2050	26,473	13,437	13,437
Validation	230	1696	230	1696
Test	684	2824	684	2824

3.3. Rust Detection Module (RDM)

The DL component of our proposed pipeline is the RDM, which performs fine-grained classification and segmentation of WRD. It consists of several essential stages, such as rust/non-rust classification, patch grouping, and semantic segmentation using the co-salient feature extraction process. All these stages are discussed in detail in the subsections below:

3.3.1. Binary Classification

The patches of the NWRD dataset are first passed through a binary classifier for the classification of each patch into rust group or non-rust group. The resolution of 224 × 224 pixels is chosen for two reasons. Firstly, the patch is large enough to ensure that sufficient context is preserved in each patch to accurately distinguish between diseased and healthy wheat leaves. Secondly, the patch is small enough to be processed more quickly by the model, as it contains fewer pixels and thus requires less computation for each forward and backward pass during training. We tested various classifiers for the binary classification task, including ensemble learning model CatBoost [33], CNN-based classifiers such as VGG16 [34], ResNet50 [35], XceptionNet [36], EfficientNet [37], and a ViT-base classifier [38].

We experimented with ViT-base [38] and demonstrated that it outperforms the ensemble learning model and the CNN models for binary rust disease classification task.

We chose the ViT-base model pre-trained on ImageNet and fine-tuned it on the training dataset to enhance its performance for classifying each patch as a rust patch or a non-rust patch. Fine-tuning involves adjusting the model's parameters using the augmented-NWRD dataset, which contains specific examples of rust and non-rust patches, making the model

more adept at this specific classification problem. For this reason, the ViT-base model has been chosen to be a part of the RDM.

3.3.2. Patch Grouping

Co-salient object detection models operate on a group of images. When multiple images are used as input, the model can extract shared features that are consistent across these images. This helps in isolating the co-salient objects that stand out in the context of the group rather than just individual images. By focusing on common features across the group, the model can better ignore outliers and emphasize the most significant objects, leading to more robust detection results. These groups of images as input provide contextual information that is not available in a single image, leveraging context across multiple images of wheat fields to help more accurately identify diseased areas of plants. It also helps co-salient object detection models to perform consistently across varying conditions, such as different lighting, angles, and backgrounds, which are common in real-world scenarios.

To create a dataset based on groups of images, we take as input the rust-based patches that are the output of the binary classifier. We select 12 patches, group them, and keep repeating this process until all patches are consumed. We selected the threshold value of 12 after experimenting with different values. Each group comprising 12 patches, with each patch of dimension 224×224 pixels, is then fed to the co-salient object detection model for the co-salient feature extraction process.

3.3.3. Co-Salient Feature Extraction

In this work, we explored salient object detection (SOD) and co-salient object detection (Co-SOD) models for segmentation of the classified rust/non-rust patches. Salient object detection focuses on identifying the most prominent or salient features in an image, those that naturally draw human attention. Salient object detection models rely on convolutional neural networks (CNNs) for global context and local spatial details of the objects to locate salient objects in an image. Ding et al. proposed the SalienDet algorithm for object detection in autonomous driving [39]. It utilizes a salience-based approach to enhance image features for generating object proposals, aiding in the detection of both known and unknown objects. An Adaptive Graph Convolution Module (AGCM) has been introduced by Lee et al. to address the limitations in salient object detection by considering image structures and pixel relations [40]. Sun et al. specify a selective feature fusion network, which enhances accuracy in salient object detection [41].

The goal of the Co-SOD detection model is to find and highlight objects that stand out (are salient) and are consistently present across a set of images. Qi et al. have presented the group-based co-saliency detection algorithm GCoNet, which is based on two important criteria of inter-group separability and intra-group compactness [42]. An improved group collaborative learning framework (GCoNet+) by the same authors is presented for efficient group-based object detection in natural scenes [43].

Since WRD segmentation is a challenging task due to the complex nature of the rust presence and backgrounds, it is difficult for the salient object detection models to accurately isolate diseased areas. Visual representation of this can be seen in Figure 3. Salient Object Detection (SOD) models can identify the leaf structures and patterns from the field images; however, localization of disease is a fine-grained task that needs a specialized solution, and co-salient object detection is a way forward in this regard.

The Co-SOD models were able to identify critical features in diseased leaf patches by contextually focusing on the prominent features present in a group of images. By leveraging this model, we effectively isolate and highlight the co-existing and salient aspects of the rust-diseased areas across the patches in a single group. The discriminative capabilities

of the model allow it to focus on the unique characteristics of NWRD, facilitating a more precise analysis of the affected patches. As we input groups of 12 patches each into the co-salient feature extractor module, it utilizes both foreground and background regions of the input images to extract features of leaves and disease. This process results in the generation of prediction maps that indicate the likelihood of rust disease presence in each patch. These prediction maps, which highlight probable diseased areas, are then stitched together based on the original configuration of patches for each image. The final output is the predicted segmentation mask, delineating the diseased regions in the images. Our experimentation with various state-of-the-art segmentation, SOD, and Co-SOD models reveals that the best feature extraction is performed by involving contextual information from a group of diseased images. Due to this reason, we selected the Co-SOD model for our RDM pipeline.

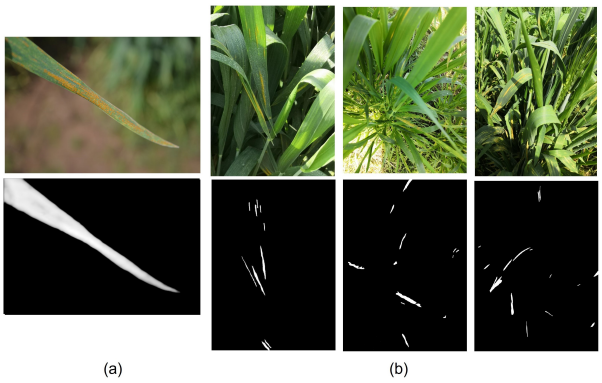


Figure 3. Visual examples of salient and co-salient object detection in a wheat field. (**a**) SOD and (**b**) Co-SOD.

3.4. Experiments

Multiple experiments have been conducted for the disease detection and segmentation process. Details of the dataset used, hardware configurations, training, and hyperparameter tuning are discussed in detail in this section.

3.4.1. Dataset

The open-source NWRD dataset published in our previous work [11] is used in this work for wheat rust segmentation. The dataset has been collected from the wheat fields of the National Agriculture Research Center (NARC) located in Islamabad, Pakistan. The multi-strain wheat crop has been utilized for real-world field dataset preparation, comprising approximately 29,000 patches of rust and non-rust data of 224 $\times$ 224 pixels. The dataset features high-resolution, densely annotated images that depict both clear and fuzzy views of WRD under natural field and weather conditions. This real-world dataset offers high-resolution images that accurately reflect the challenges of diseased crop fields. Each image includes multiple diseased leaves at various stages of rust infection, providing a comprehensive representation of the disease's progression.

Table 1 gives details of the total number of patches of the NWRD dataset and the augmented-NWRD dataset utilized for our classifier and co-saliency model training.

3.4.2. Implementation Details

All experiments were conducted on a system equipped with an Nvidia RTX A5000 GPU sourced from Kaiserslautern, Germany , running the Linux operating system. The setup utilized the PyTorch framework within the Anaconda environment, with Python 3.8 and CUDA 11.7. The hardware configuration included an Intel(R) Core i7-9700 CPU @ 3.00 GHz (8 cores) and 24 GB of memory.

3.5. RDM Training

As the first step, the RDM classifies input patches into rust and non-rust patches. As the second step, the rust patches are grouped and processed for disease detection and segmentation by the Co-SOD model to produce the final prediction maps, which are concatenated to form the predicted masks. For the classification step, we performed two sets of experiments to evaluate the efficacy of classifiers in the binary wheat rust classification task. Initially, we trained the classifiers using the NWRD dataset, employing 2050 rust patches and 2050 non-rust patches selected from 26,473 non-rust patches. By balancing the number of rust and non-rust patches in training, we aimed to prevent bias towards the non-rust class. Subsequently, we investigated the impact of data augmentation by retraining the classifier on the augmented-NWRD dataset.

As already elaborated, we conducted experiments on several CNN-based classifiers, the ensemble learning model, and a transformer-based classifier to identify the best model for wheat rust disease classification. The ViT-base model, pre-trained on ImageNet and fine-tuned on the NWRD dataset, achieved the highest performance for the wheat rust classification task. The training process utilized the SGD optimizer with an initial learning rate of 3×10^{-8}, a weight decay of 0.001, and a batch size of 8, over 50 epochs, as outlined in Table 2, with binary cross-entropy loss to optimize the classifier's performance. For the ensemble learning model CatBoost, the learning rate was set to 0.1, the boosting type was plain and cross-entropy loss was used as the loss function.

Table 2. Hyperparameters used for the training of our RDM for the task of classification and segmentation.

Sr. No.	Parameters	Classification	Segmentation
1.	Epochs	50	100
2.	Batch Size	8	1
3.	Optimizer	SGD	Adam
4.	Evaluation Metrics	F1 score, Accuracy	F1 Score, Mean IoU
5.	Loss Function	Cross-entropy	IoU loss, Self-contrastive loss

For the co-salient feature extraction process of WRD, we utilized three different co-salient object detection models, including GCoNet [42], GCoNet+ [43], and DCFM [13]. We fine-tuned all these pre-trained models on the NWRD dataset for rust disease detection and segmentation. Two sets of training were conducted for tailoring the Co-SOD models for the rust disease detection problem, one with the NWRD dataset and another one with the augmented-NWRD dataset.

For all Co-SOD model training, the Adam optimizer with an initial learning rate of 1×10^{-4}, $\beta_1 = 0.9$, and $\beta_2 = 0.99$ values was used to train the models for 100 epochs. Computations on each group of images were carried out individually, i.e., 12 patches per group with a batch-size of 1. The Intersection over Union (IoU) and self-contrastive loss were combined for this training as given by the authors in [13].

When we combine ViT-base and the democratic co-salient feature-mining model (DCFM) pre-trained on the COCO dataset with CoCA as the validation dataset by Yu et al. [13] and fine-tuned on the augmented-NWRD dataset, the best results for the segmentation of diseased leaves takes place. The feature maps generated by our proposed technique can be seen in Figure 4. Therefore, this tailored Co-SOD model was made part of the RDM as the co-salient feature extractor in our proposed WRD pipeline. The inference steps for the wheat rust disease detection pipeline are given as pseudo-code in Algorithm 1.

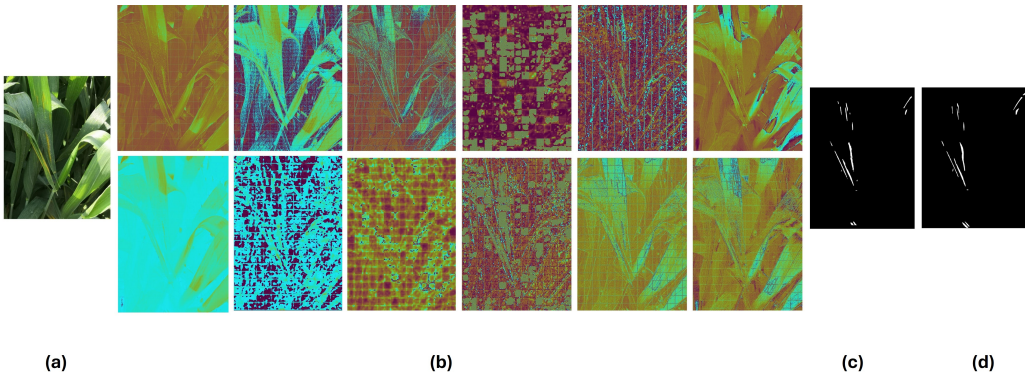


Figure 4. Feature maps generated by our proposed technique for wheat rust segmentation that illustrate disease detection across various layers of the model: (**a**) input image, (**b**) feature maps of internal layers of the model, (**c**) predicted mask, and (**d**) ground truth.

Algorithm 1 Rust_Segmentation_Inference_Pipeline

procedure RUST_SEGMENTATION_INFERENCE_PIPELINE
  Input: Input image I
  Output: Final prediction mask M
  ▷ Step 1: Divide the image into patches
  Patches(I_p) ← DivideIntoPatches(I)
  ▷ Step 2: Use ViT to classify rust patches
  RustPatches(P_r) ← []
  for each patch in I_p **do**
    if ViT_Classify(patch) == "rust" **then**
      RustPatches.append(patch)
    end if
  end for
  ▷ Step 3: Make patch groups, each of 12 rust patches
  GroupedPatches(G_p) ← GroupPatches$(P_r, 12)$
  ▷ Step 4: Pass groups to Co-SOD model for segmentation
  SegmentedGroups(G_s) ← []
  for each group in G_p **do**
    SegmentedGroup ← Co-SOD_Segment$(group)$
    SegmentedGroups.append(SegmentedGroup)
  end for
  ▷ Step 5: Concatenate patches to form the final prediction mask
  M ← ConcatenatePatches(G_s)
  return M
end procedure

3.6. Evaluation Metrics

The effectiveness of the proposed RDM was evaluated and compared against state-of-the-art models using the following metrics:

1. Intersection over Union (IoU): IoU is a standard metric for segmentation tasks that measures the degree of overlap between the predicted segmentation mask and the ground truth. It is expressed as the ratio of the intersecting area to the combined area (union) of the predicted and ground truth masks. Higher IoU values reflect improved segmentation performance.
2. F1 score: This metric is based on the precision and recall scores and is particularly useful when the data are imbalanced.
3. Accuracy: This metric represents the ratio of instances classified correctly to the total instances present.

4. Model Complexity: To assess model complexity, we employ two key parameters: training time and Giga Floating-Point Operations (GFLOPs).
   When performing segmentation on high-resolution images, the necessity to process each pixel individually for segmentation significantly increases the training time. Therefore, to better understand the model's complexity, we use training time as a key parameter instead of number of parameters. By analyzing the training duration, we can gain insights into the computational and environmental demands of different segmentation models, ultimately aiding in the selection of an efficient model. A lower training time indicates higher efficiency regarding computational resources and environmental impacts.
   The GFLOPs metric is used to measure the computational load required to execute a model. Specifically, this indicates how many operations are needed to run a single instance of a given model. For Pytorch models, the open-source tool PTFLOPS [44] has been used to calculate the GFLOPs of models. However, for other models such as SegFormer and Yolov8-Seg, THOP, a pytorch-based API, has been used to calculate GFLOPs [45].

4. Results

The experimental results for various models and our proposed pipeline have been presented in the following section.

4.1. Quantitative and Qualitative Results

Figure 5 illustrates the test F1 scores of the NWRD and augmented-NWRD datasets on the CNN and ViT-base classifiers. The results show that the model performs much better on the augmented-NWRD dataset due to balanced rust and non-rust patches in the training set. Results of all classifiers are consistent with the fact that augmentations diversify training data, aiding models in learning invariant features and mitigating over-fitting, thus improving the overall F1 score of the classifier. The ViT-base classifier outperforms other classifiers using the augmented-NWRD dataset.

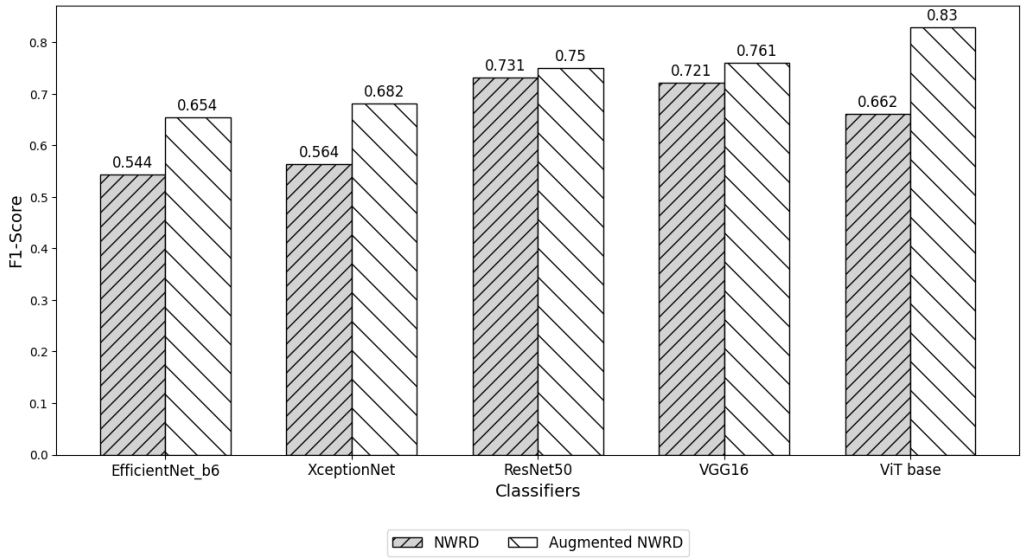


Figure 5. Result comparison of different classifiers on the NWRD and augmented-NWRD datasets for binary classification. ViT-base achieves the highest F1 score among the other classifiers.

Table 3 provides a detailed comparison between CNN-based and transformer-based classifiers for binary classification of rust/non-rust patches on the augmented-NWRD

dataset. The table elaborates that ViT-base outperforms other classification models by achieving 0.879 precision, 0.830 F1 score, and 93.73% accuracy.

Table 3. Result comparison of state-of-the-art classifiers for classification on augmented-NWRD dataset using performance metrics of precision, recall, F1 score, and accuracy. ViT-base outperforms other classifiers by achieving an F1 score of 0.83.

Classifier	True Positive	False Positive	True Negative	False Negative	Precision	Recall	F1 score	Accuracy (%)
CatBoost	505	352	2472	179	0.59	0.74	0.66	85.00
EfficientNet_b6	392	122	2702	292	0.763	0.573	0.654	88.19
XceptionNet	418	123	2701	266	0.772	0.611	0.682	89.01
ResNet50	500	150	2674	184	0.769	0.731	0.750	90.47
VGG16	515	154	2670	169	0.770	0.753	0.761	90.79
ViT-base	538	74	2750	146	0.879	0.787	0.830	93.73

For the segmentation of the augmented-NWRD dataset, our proposed RDM achieved an F1 score of 0.638 and IoU of 0.470, outperforming the UNet segmentation model, as shown in Table 4. Furthermore, among Co-SOD models, DCFM outperforms GCoNet and GCoNet+ on the augmented-NWRD dataset for WRD and therefore is selected to be part of our proposed RDM. Furthermore, with the RDM, the training time is 5× less as compared to our previous work.

Table 4. Comparison of various segmentation models on the NWRD dataset using the metrics of F1 score, IoU, training time, and GFLOPs, where ↑ and ↓ denote larger and smaller being better, respectively. Our proposed RDM, which is based on Vit-base and DCFM, yields the best results with an F1 score of 0.638 and IoU of 0.471.

Sr. No.	Technique	Model	F1 Score ↑	IoU ↑	Training Time (min) ↓	GFLOPs ↓
1	SOD [46]	ABiU-NeT	0.453	0.396	1863	17.02
2	CNN [47]	YOLOv8n-Seg	0.421	0.458	110	1.483
3	Transformer [23]	SegFormer_MiT_B5	0.550	0.300	1900	16.94
4	CNN [11]	UNet with APF	0.557	0.438	4791	83.86
5	Classifier+Co-SOD (This Work)	ViT_base + GCoNet	0.564	0.395	1094	96.24
6	Classifier+Co-SOD (This Work)	ViT_base + GCoNet+	0.579	0.413	1318	106.88
7	RDM(This Work)	ViT_base + DCFM	0.638	0.471	921	97.34

4.2. Comparison with State-of-the-Art Methods

The most recent research work on wheat rust segmentation is presented by Deng et al. [23]. In their study, the authors used Segformer to segment areas affected by wheat stripe rust. To address the class imbalance in their proprietary wheat stripe rust dataset, data augmentation techniques were utilized. As their dataset is not available for experimentation, we applied the SegFormer model to the augmented-NWRD for WRD segmentation, ensuring a fair comparison with our proposed model.

Table 4 shows a quantitative comparison of our proposed RDM against the previous works such as the UNet [11], SegFormer [23], ABiU-NeT [46], and YOLOv8n-Seg [47] models on the augmented-NWRD dataset. The RDM outperforms the previous works in terms of F1 score, Intersection over Union (IoU), and training time.

Although the RDM (proposed model) has higher GFLOPs compared to previous segmentation-based models, it uses a classifier at the initial stage. This classifier selectively identifies only a small number of patches containing rust for segmentation. Consequently, the training time of the proposed RDM pipeline is significantly reduced. Alternatively, in previous works, the experiments (numbered 1 to 4 in Table 4) perform pixel-wise segmentation on the entire dataset. Due to this, despite the lower GFLOPs, the training

time is much higher. It is important to mention that GFLOPs are calculated on an input image size of 224 × 224 for all the experiments. By using a classifier to localize the rust areas requiring segmentation, the RDM model enhances efficiency while maintaining a high F1 score, demonstrating a notable improvement over previous segmentation-only approaches.

The underlying architectures of all the segmentation models in Table 4 (CNN, transformer, SOD, Co-SOD) are inherently different, which directly impacts their computational requirements. These differences arise from variations in topology, such as the number of layers, type of operations (e.g., convolutions, attention mechanisms), and model complexity. As a result, the training time and GFLOP metrics vary significantly between these models.

Figure 6 illustrates a visual comparison of the augmented-NWRD dataset, the ground truth masks, and the predicted masks generated by our proposed RDM and the UNet, SOD-based AbiU-Net, SegFormer, and YOLOseg models.

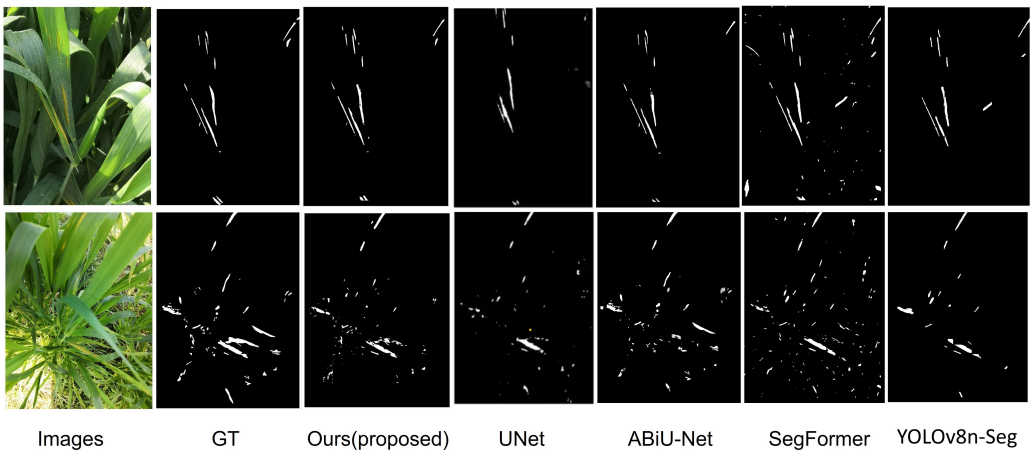


Figure 6. Qualitative comparison of our proposed wheat rust disease detection solution with state-of-the-art methods. Visual results show that the classifier and co-salient feature extractor method performs the most accurate segmentation among all.

5. Discussion

Early segmentation of WRD is critical for safeguarding global food security and ensuring sustainable agricultural practices. WRD, a fungal disease that threatens wheat crops worldwide, results in yield losses if not identified and managed promptly. By leveraging advanced AI-based machine learning algorithms and image processing techniques, we can accurately and rapidly identify early signs of WRD in large-scale farming operations. This enables timely intervention and targeted treatment, resulting in reducing the spread of the disease and minimizing crop damage. The literature highlights research work in the field of disease detection focusing either on disease classification or on semantic segmentation techniques for disease detection. MobileViT has also been utilized for wheat disease identification and application on smart devices for field-based disease detection. However, the input data in most cases include focused images of disease. In that scenario, UNet, DeepLabV3, the Segment Anything Model (SAM) SegFormer, MobileVit, or a similar model can achieve good results. Conversely, when early-stage disease detection is required, where the diseased region is minimal compared to its surroundings, semantic segmentation models often fall short of delivering reliable outcomes. This limitation arises due to the distinct field conditions during the early stages of the disease, which differ significantly from the conditions when the disease is more widespread. To this end, our work in this research work addresses the problem of automatically detecting WRD in the real-world NWRD field dataset. Notably, this dataset presents a significant challenge, as diseased areas account for only about 10% of the total coverage, with the majority consisting of healthy

wheat leaves. This dataset comprises high-resolution WRD infected images, featuring diverse illumination conditions and intricate backgrounds. For segmentation of WRD, we employed the Co-SOD technique to develop a disease detection pipeline that is capable of performing segmentation of disease in intricate scenes. Extensive experimentation proves that the proposed pipeline improves the accuracy of WRD segmentation as compared to the previous works while also taking 5× less time to train.

The Co-SOD approach for disease detection and segmentation has not been investigated for crop diseases in general and WRD in particular. The advantage of using the co-salient feature extraction technique for crop disease detection is its ability to incorporate both local and global context during the feature extraction process. This shows an increase in sensitivity of the disease segmentation pipeline, enabling the detection of subtle differences between healthy and infected crops, which is crucial for early-stage segmentation.

This advancement underscores the potential of co-saliency techniques in enhancing precision agriculture, enabling timely and accurate identification of disease-affected areas, ultimately contributing to improved crop productivity and food security. Our findings pave the way for further exploration and refinement of co-saliency applications in the agricultural contexts, promising significant benefits for farmers and the agricultural industry at large. In future work, we plan to extend the application of the proposed technique to a variety of crop diseases to evaluate its robustness and generalizability. This will involve fine-tuning the model on datasets from different crops and disease types to assess its adaptability and performance in diverse scenarios.

6. Conclusions

In conclusion, our research demonstrates a novel application of the Co-SOD model in the field of agricultural disease management, specifically targeting the detection and segmentation of WRD. Utilizing a two-stage RDM model, this work tackles the intricate and varied spread of WRD in the NWRD dataset. For the first stage of the RDM, which requires binary classification, we evaluated multiple state-of-the-art classifiers. Of these classifiers, ViT outperformed others by achieving the highest F1 score (0.83). The second stage involves a co-salient object detection model segmenting rust in the classified rust patches of the NWRD dataset. This approach shows a notable improvement in the segmentation performance by achieving an F1 score of 0.638 with 5× less training time as compared to the previous works. In future research, we intend to enhance this study by implementing a unified model as compared to the current two-stage model for the segmentation of WRD and investigate other crop diseases to evaluate its robustness.

Author Contributions: Conceptualization, H.A., F.S., M.J.K. and M.M.G.; methodology, H.A., M.M.G. and H.M.; data acquisition, H.A. and M.A.A.; model implementation, H.M., M.A.A. and M.M.G.; validation, H.M., M.A.A., M.M.G. and F.S.; writing—original draft preparation, H.A., H.M. and M.A.A.; writing—review and editing, F.S., M.J.K., M.M.G., C.W. and N.W.; visualization, F.S., M.M.G., C.W. and N.W.; supervision, F.S., C.W., N.W. and M.J.K. All authors have read and agreed to the published version of the manuscript.

Funding: The research reported in this work was partially supported by Carl Zeiss Stiftung, Germany, under the Sustainable Embedded AI project (P2021-02-009).

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The list of abbreviations is given below:

WRD	Wheat Stripe Rust Disease
Co-SOD	Co-Salient Object Detection
SOD	Salient Object Detection
RDM	Rust Detection Module
ML	Machine Learning
DL	Deep Learning
ViT	Vision Transformer
CNN	Convolutional Neural Network

References

1. Kolmer, J.A.; Fajolu, O. Virulence phenotypes of the wheat leaf rust pathogen, Puccinia triticina, in the United States from 2018 to 2020. *Plant Dis.* **2022**, *106*, 1723–1729. [CrossRef] [PubMed]
2. Figueroa, M.; Hammond-Kosack, K.E.; Solomon, P.S. A review of wheat diseases—A field perspective. *Mol. Plant Pathol.* **2018**, *19*, 1523–1536. [CrossRef] [PubMed]
3. Roelfs, A.P.; Singh, R.P.; Saari, E. *Rust Diseases of Wheat: Concepts and Methods of Disease Management*; Cimmyt: Mexico City, Mexico, 1992.
4. Wellings, C.R. Global status of stripe rust: A review of historical and current threats. *Euphytica* **2011**, *179*, 129–141. [CrossRef]
5. Huerta-Espino, J.; Singh, R.; Crespo-Herrera, L.A.; Villaseñor-Mir, H.E.; Rodriguez-Garcia, M.F.; Dreisigacker, S.; Barcenas-Santana, D.; Lagudah, E. Adult Plant Slow Rusting Genes Confer High Levels of Resistance to Rusts in Bread Wheat Cultivars From Mexico. *Front. Plant Sci.* **2020**, *11*, 824. [CrossRef]
6. Shafi, U.; Mumtaz, R.; Haq, I.U.; Hafeez, M.; Iqbal, N.; Shaukat, A.; Zaidi, S.M.H.; Mahmood, Z. Wheat yellow rust disease infection type classification using texture features. *Sensors* **2021**, *22*, 146. [CrossRef]
7. Shafi, U.; Mumtaz, R.; Qureshi, M.D.M.; Mahmood, Z.; Tanveer, S.K.; Haq, I.U.; Zaidi, S.M.H. Embedded AI for wheat yellow rust infection type classification. *IEEE Access* **2023**, *11*, 23726–23738. [CrossRef]
8. Fang, X.; Zhen, T.; Li, Z. Lightweight multiscale CNN model for wheat disease detection. *Appl. Sci.* **2023**, *13*, 5801. [CrossRef]
9. Pan, Q.; Gao, M.; Wu, P.; Yan, J.; AbdelRahman, M.A. Image classification of wheat rust based on ensemble learning. *Sensors* **2022**, *22*, 6047. [CrossRef]
10. Hayıt, T.; Erbay, H.; Varçın, F.; Hayıt, F.; Akci, N. The classification of wheat yellow rust disease based on a combination of textural and deep features. *Multimed. Tools Appl.* **2023**, *82*, 47405–47423. [CrossRef]
11. Anwar, H.; Khan, S.U.; Ghaffar, M.M.; Fayyaz, M.; Khan, M.J.; Weis, C.; Wehn, N.; Shafait, F. The NWRD Dataset: An Open-Source Annotated Segmentation Dataset of Diseased Wheat Crop. *Sensors* **2023**, *23*, 6942. [CrossRef]
12. Patterson, D.; Gonzalez, J.; Le, Q.; Liang, C.; Munguia, L.M.; Rothchild, D.; So, D.; Texier, M.; Dean, J. Carbon emissions and large neural network training. *arXiv* **2021**, arXiv:2104.10350.
13. Yu, S.; Xiao, J.; Zhang, B.; Lim, E.G. Democracy does matter: Comprehensive feature mining for co-salient object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 979–988.
14. Fan, D.P.; Li, T.; Lin, Z.; Ji, G.P.; Zhang, D.; Cheng, M.M.; Fu, H.; Shen, J. Re-thinking co-salient object detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 4339–4354. [CrossRef] [PubMed]
15. Li, Z.; Fang, X.; Zhen, T.; Zhu, Y. Detection of wheat yellow rust disease severity based on improved GhostNetV2. *Appl. Sci.* **2023**, *13*, 9987. [CrossRef]
16. Niu, X.; Wang, M.; Chen, X.; Guo, S.; Zhang, H.; He, D. Image segmentation algorithm for disease detection of wheat leaves. In Proceedings of the IEEE International Conference on Advanced Mechatronic Systems, Kumamoto, Japan, 10–12 August 2014; pp. 270–273.
17. Bukhari, H.R.; Mumtaz, R.; Inayat, S.; Shafi, U.; Haq, I.U.; Zaidi, S.M.H.; Hafeez, M. Assessing the impact of segmentation on wheat stripe rust disease classification using computer vision and deep learning. *IEEE Access* **2021**, *9*, 164986–165004. [CrossRef]
18. Zhang, D.; Gu, C.; Wang, Z.; Zhou, X.; Li, W. Evaluating the efficacy of fungicides for wheat scab control by combined image processing technologies. *Biosyst. Eng.* **2021**, *211*, 230–246. [CrossRef]
19. Li, Y.; Qiao, T.; Leng, W.; Jiao, W.; Luo, J.; Lv, Y.; Tong, Y.; Mei, X.; Li, H.; Hu, Q.; et al. Semantic Segmentation of Wheat Stripe Rust Images Using Deep Learning. *Agronomy* **2022**, *12*, 2933. [CrossRef]
20. Li, G.; Wang, Y.; Zhao, Q.; Yuan, P.; Chang, B. PMVT: A lightweight vision transformer for plant disease identification on mobile devices. *Front. Plant Sci.* **2023**, *14*, 1256773. [CrossRef]

21. Singh, A.K.; Rao, A.; Chattopadhyay, P.; Maurya, R.; Singh, L. Effective plant disease diagnosis using Vision Transformer trained with leafy-generative adversarial network-generated images. *Expert Syst. Appl.* **2024**, *254*, 124387. [CrossRef]
22. Zhang, K.; Ma, L.; Cui, B.; Li, X.; Zhang, B.; Xie, N. Visual large language model for wheat disease diagnosis in the wild. *Comput. Electron. Agric.* **2024**, *227*, 109587. [CrossRef]
23. Deng, J.; Lv, X.; Yang, L.; Zhao, B.; Zhou, C.; Yang, Z.; Jiang, J.; Ning, N.; Zhang, J.; Shi, J.; et al. Assessing macro disease index of wheat stripe rust based on segformer with complex background in the field. *Sensors* **2022**, *22*, 5676. [CrossRef]
24. Li, W.; Yu, X.; Chen, C.; Gong, Q. Identification and localization of grape diseased leaf images captured by UAV based on CNN. *Comput. Electron. Agric.* **2023**, *214*, 108277. [CrossRef]
25. Kerkech, M.; Hafiane, A.; Canals, R. Vine disease detection in UAV multispectral images using optimized image registration and deep learning segmentation approach. *Comput. Electron. Agric.* **2020**, *174*, 105446. [CrossRef]
26. Wang, Z.; Zhang, S. Segmentation of corn leaf disease based on fully convolution neural network. *Acad. J. Comput. Inf. Sci.* **2018**, *1*, 9–18.
27. Wang, X.f.; Wang, Z.; Zhang, S.w. Segmenting crop disease leaf image by modified fully-convolutional networks. In Proceedings of the Intelligent Computing Theories and Application: 15th International Conference, ICIC 2019, Nanchang, China, 3–6 August 2019; Proceedings, Part I 15; Springer: Cham, Switzerland, 2019; pp. 646–652.
28. Lin, K.; Gong, L.; Huang, Y.; Liu, C.; Pan, J. Deep learning-based segmentation and quantification of cucumber powdery mildew using convolutional neural network. *Front. Plant Sci.* **2019**, *10*, 155. [CrossRef]
29. Stewart, E.L.; Wiesner-Hanks, T.; Kaczmar, N.; DeChant, C.; Wu, H.; Lipson, H.; Nelson, R.J.; Gore, M.A. Quantitative phenotyping of northern leaf blight in UAV images using deep learning. *Remote Sens.* **2019**, *11*, 2209. [CrossRef]
30. Wang, Q.; Qi, F.; Sun, M.; Qu, J.; Xue, J. Identification of tomato disease types and detection of infected areas based on deep convolutional neural networks and object detection techniques. *Comput. Intell. Neurosci.* **2019**, *2019*, 9142753. [CrossRef]
31. Badrinarayanan, V.; Kendall, A.; Cipolla, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [CrossRef]
32. Fischer, S.M.; Felsner, L.; Osuala, R.; Kiechle, J.; Lang, D.M.; Peeken, J.C.; Schnabel, J.A. Progressive Growing of Patch Size: Resource-Efficient Curriculum Learning for Dense Prediction Tasks. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Marrakesh, Morocco, 6–10 October 2024; Springer: Berlin/Heidelberg, Germany, 2024; pp. 510–520.
33. Dorogush, A.V.; Gulin, A.; Gusev, G.; Kazeev, N.; Prokhorenkova, L.O.; Vorobev, A. Fighting biases with dynamic boosting. *arXiv* **2017**, arXiv:1706.09516.
34. Simonyan, K. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556
35. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
36. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.
37. Tan, M.; Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In Proceedings of the International Conference on Machine Learning, PMLR, Long Beach, CA, USA, 9–15 June 2019; pp. 6105–6114.
38. Alexey, D. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929.
39. Ding, N.; Zhang, C.; Eskandarian, A. SalienDet: A saliency-based feature enhancement algorithm for object detection for autonomous driving. *IEEE Trans. Intell. Veh.* **2023**, *9*, 2624–2635. [CrossRef]
40. Lee, Y.; Lee, M.; Cho, S.; Lee, S. Adaptive Graph Convolution Module for Salient Object Detection. In Proceedings of the IEEE International Conference on Image Processing (ICIP), Kuala Lumpur, Malaysia, 8–11 October 2023; pp. 1395–1399.
41. Sun, F.; Yuan, X.; Zhao, C. Selective feature fusion network for salient object detection. *IET Comput. Vis.* **2023**, *17*, 483–495. [CrossRef]
42. Fan, Q.; Fan, D.P.; Fu, H.; Tang, C.K.; Shao, L.; Tai, Y.W. Group collaborative learning for co-salient object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 12288–12298.
43. Zheng, P.; Fu, H.; Fan, D.P.; Fan, Q.; Qin, J.; Tai, Y.W.; Tang, C.K.; Van Gool, L. Gconet+: A stronger group collaborative co-salient object detector. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 10929–10946. [CrossRef]
44. Sovrasov, V. ptflops: A FLOPS Counting Tool for Neural Networks in the PyTorch Framework. 2018. Available online: https://github.com/sovrasov/flops-counter.pytorch (accessed on 14 January 2025).
45. Pytorch. THOP:PyTorch-OpCounter. Available online: https://github.com/Lyken17/pytorch-OpCounter (accessed on 14 January 2025).

46. Qiu, Y.; Liu, Y.; Zhang, L.; Lu, H.; Xu, J. Boosting salient object detection with transformer-based asymmetric bilateral U-Net. *IEEE Trans. Circuits Syst. Video Technol.* **2023**, *34*, 2332–2345. [CrossRef]
47. Varghese, R.; M., S. YOLOv8: A Novel Object Detection Algorithm with Enhanced Performance and Robustness. In Proceedings of the 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS), Chennai, India, 18–19 April 2024; pp. 1–6. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.