Aerial Visual Perception in Smart Farming: Field Study of Wheat Yellow Rust Monitoring

Jinya Su, Dewei Yi, Baofeng Su, Zhiwen Mi, Cunjia Liu, Xiaoping Hu, Xiangming Xu, Lei Guo and Wen-Hua Chen, *Fellow, IEEE*

***Abstract*—Agriculture is facing severe challenges from crop stresses, threatening its sustainable development and food security. This work exploits aerial visual perception for yellow rust disease monitoring, which seamlessly integrates state-of-the-art techniques and algorithms including UAV sensing, multispectral imaging, vegetation segmentation and deep learning U-Net. A field experiment is designed by infecting winter wheat with yellow rust inoculum, on top of which multispectral aerial images are captured by DJI Matrice 100 equipped with RedEdge camera. After image calibration and stitching, multispectral orthomosaic is labelled for system evaluation by inspecting high-resolution RGB images taken by Parrot Anafi Drone. The merits of the developed framework drawing spectral-spatial information concurrently are demonstrated by showing improved performance over purely spectral based classifier by the classical random forest algorithm. Moreover, various network input band combinations are tested including three RGB bands and five selected spectral vegetation indices by Sequential Forward Selection strategy of Wrapper algorithm.**

***Index Terms*—Deep learning; Multispectral image; Precision agriculture; Semantic segmentation; U-Net; Unmanned Aerial Vehicle (UAV).**

I. Introduction

Visual perception is to interpret the environment by the light (in the form of images captured by various cameras) reflected by the objects via image analysis [1] and is now finding a wide range of applications in smart society (e.g. transportation surveillance [2], aircraft detection [3], smart health [4], industrial inspection [5]). Following this line of thought, this work aims to exploit aerial visual perception in smart farming to tackle the grand challenge facing modern agriculture: feeding a growing world population with an ageing structure while protecting the environment. This is achieved by developing a disease monitoring framework for precision stress management. To this end, this work proposes an automated monitoring framework for yellow rust disease in winter wheat by seamlessly integrating deep learning algorithms and multispectral aerial images collected by a small Unmanned Aerial Vehicle (UAV) at an experimental wheat field.

Wheat is the most widely grown crop in the world, providing 20% of protein and food calories for 4.5B people. Its demand is also increasing with a growing world population (60% more by 2050 with a predicted population of 9B). However, wheat production is now facing a number of challenges from abiotic stresses, pathogens and pests due to climate changes. Among them, wheat yellow (or stripe) rust, caused by *Puccinia striiformis* f. sp. *tritici* (Pst), is a devastating wheat disease worldwide, particularly in regions with temperate climates [6]. This disease develops and spreads very quickly under favourable environmental conditions such as a temperature of 5-24^oC, a moderate precipitation in spring. It is estimated that yield loss caused by yellow rust disease is at least 5.5 million tons per year at a global level.

An accurate and timely monitoring of yellow rust disease plays a paramount role in its precision management, paving the way for sustainable crop production and food security [7]. In particular, the disease mapping enables a timely and precise fungicide application so that its adverse effects can be effectively minimised with a reduced use of pesticides compared to conventional uniform spraying strategy. Besides, the automated disease monitoring system can also help breeders in selecting suitable wheat genotypes that are resistant to yellow rust disease in breeding programmes. Rust disease usually leads to some physical and chemical changes on wheat leaves including ChlorophyII content reduction, water loss, and visual rust symptoms (i.e. yellow-orange to reddish-brown spores). These changes can be effectively captured by spectral sensors (e.g. optical cameras) or human eyes. The current practice of disease monitoring relies on visual inspection via naked eyes [8]. This approach is accurate, however, is time-consuming, labour-intensive, costly and not suitable for applications at field scales [7] due to a large number of required sampling points. Therefore, there is a trend to adopt imaging approaches for an automated crop disease monitoring in recent years.

Various types of sensors have been investigated for disease monitoring in the literature: from low-cost RGB visual [9] to high-cost hyperspectral camera [6] and from ground proximity sensing [10] to aircraft (or even satellite) remote sensing [7].

Manuscript received... This work was supported by Science and Technology Facilities Council (STFC) under Newton fund with Grant No. ST/N006852/1. (Corresponding authors: Cunjia Liu and Xiaoping Hu).
Jinya Su is with School of Computer Science and Electronic Engineering, University of Essex, Colchester, CO4 3SQ, UK.
Dewei Yi is with Department of Computing Science, University of Aberdeen, Aberdeen, AB24 3FX, UK.
Baofeng Su and Zhiwen Mi are with College of Mechanical and Electronic Engineering, Northwest A&F University, Yangling, Shaanxi 712100, China.
Cunjia Liu and Wen-Hua Chen are with Department of Aeronautical and Automotive Engineering, Loughborough University, Loughborough LE11 3TU, UK. E-mail: C.Liu5@lboro.ac.uk;
Xiaoping Hu is with State Key Laboratory of Crop Stress Biology for Arid Areas, College of Plant Protection, Northwest A&F University, Yangling, Shaanxi 712100, China. E-mail: xphu@nwsuaf.edu.cn;
Xiangming Xu is with Department of Pest & Pathogen Ecology, NIAB EMR, West Malling ME19 6BJ, UK.
Lei Guo is with School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China.

In particular, among various sensing platforms, UAV remote sensing with a user-defined spatial-temporal resolution, a low cost and a high flexibility is drawing increasing popularity for applications at farmland scales and has been applied widely since 2010 in many areas such as disease monitoring [6], weed mapping [11, 12], and stress detection.

There are also several studies on UAV remote sensing for yellow rust disease monitoring. RGB image is adopted in [9] at an altitude of 100m, which shows that Red is the most informative visible band. Five-bands multispectral image is applied in [7, 8] at an altitude of about 20m; it is shown that Red and NIR bands are most effective and their normalized difference NDVI results in even better performance. Hyperspectral imaging is also used in [6] at a flight height of 30 m, where the problem of yellow rust monitoring is cast as an image level (3D image block) classification and is solved by advanced Convolutional Neural Network (CNN) classifier. Multispectral camera is used in this study since it has visible-NIR bands and is easy to operate.

Semantic segmentation, different from image level classification (generating only one label for the whole input image), is to classify input image into a number of class labels for each pixel. This technique is especially preferred in applications such as remote sensing [13] and biomedical image analysis [14]. Traditional ways for semantic segmentation include point, line and edge detection methods, thresholding, region-based, pixel-based clustering and morphological approaches. Recently, the challenging crop stress monitoring task is also formulated as a semantic segmentation problem and addressed by using CNN due to its strong ability in automatically extracting spectral-spatial features. For instance, the so-called Pixel-based CNN is applied in [15] for satellite image classification, where the class label at each pixel is derived by classifying the neighbouring patch centred at the pixel by CNN. To avoid selecting a suitable patch size and reduce the computation load, a Fully Convolutional Network (FCN) is applied in [11] for weed mapping by using RGB aerial image and is shown to outperform the Pixel-level CNN. The encoder-decoder cascaded CNN, SegNet, is also applied in [12] for weed mapping by using multispectral image. Very recently, the state-of-the-art U-Net is applied in [10] for leaf level disease segmentation of cucumber leaf with promising performance. U-Net and mask R-CNN [16] are compared [13] for tree canopy segmentation by using UAV RGB image at 30m. To summarize, the following observations are presented to motivate the research in this study:

(i) RGB image only possesses three visible bands (Blue, Green and Red), and its image quality is easily susceptible to environmental variations [9] in comparison to multispectral image with an accurate calibration panel;
(ii) Disease monitoring based on purely spectral information [7], may lead to a high proportion of false positives due to the spatial inhomogeneity;
(iii) Pixel-level CNN is effective in extracting spectral-spatial features [6], [15], however, patch size is empirically determined and it also involves a high computation load;
(iv) Semantic segmentation based on FCN (e.g. FCN-8 [11], SegNet [12], U-Net [16]), is proved to be effective in a number of crop stress monitoring, however, its performance is yet to be assessed for yellow rust disease.

Therefore, this work aims to develop an automated yellow rust disease monitoring framework by integrating UAV remote sensing, multispectral imaging, and deep learning U-Net algorithm. The developed framework is initially validated by real-life field experiments with promising performance, where aerial images and ground data are collected on wheat field infected by yellow rust disease. To the best of the authors' knowledge, this work is the first attempt to integrate deep learning U-Net, UAV multispectral and RGB images to address the problem of wheat yellow rust monitoring. To be more precise, the main contributions are summarized as below.

(1) *A wheat yellow rust monitoring framework is proposed by integrating UAV remote sensing, multispectral imaging and U-Net deep learning network;*
(2) *The advantages of using all five spectral bands are tested against only using three RGB bands and selected SVIs;*
(3) *Deep learning algorithms are tested against spectral based classifier by the classical random forest algorithm;*
(4) *Field experiments are designed to generate an open-access dataset, against which the developed framework is initially validated with promising performance.*

II. Experiment design

A. Field experiment setup

Field experiments are carried out in 2019 at Caoxinzhuang experimental station of Northwest Agriculture and Forestry (A&F) University, Yangling, Shanxi Province, China. Some background information about this region such as geographic location, soil property and climate is referred to [8]. Xiaoyan 22, one wheat variety susceptible to yellow rust disease, is chosen. In order to inoculate wheat plants with yellow rust inoculum, the mixed Pst races (CYR 29/30/31/32/33) are applied to wheat seedlings in March/2019 by using the approach described in [9]. As displayed in Fig 1 (letters A, B, and C denote three replicates; numbers 0-5 represent different levels of yellow rust inoculum), each plot ($2m \times 2m$) in each replicate is randomly inoculated with one of the six levels of yellow rust inoculum: 0g (health wheat plots for blank comparison), 0.15g, 0.30g, 0.45g, 0.6g and 0.75g respectively. The 18 wheat plots are well separated from each other by healthy wheat to minimise disease cross-infection.

B. Multispectral imaging and data pre-processing

In this study, a commercial off-the-shelf DJI Matrice 100 (M100) Quadcopter (DJI Company, Shenzhen, China) and a five-band multispectral visible-infrared camera (RedEdge, MicaSense Inc., Seattle, USA) (see [8] for its specifications such as weight, dimensions, image size) are integrated to be the UAV imaging platform for yellow rust disease monitoring. The flowchart of aerial imaging and image pre-processing steps is displayed in Fig 2.

Data collection is done on 02/May/2019, when yellow rust symptoms are visible as shown in Fig 3. The UAV flight height is set to be 20m above ground, where the image

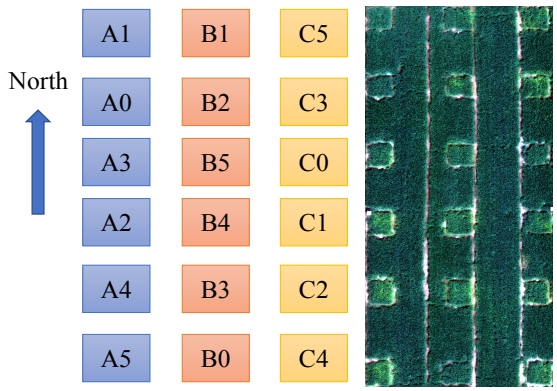

Fig. 1. Layout of wheat yellow rust disease experiment: three replicates (column-wise) with various levels of yellow rust inoculum (left); false-color RGB image of the wheat field at diseased stage on 02/May/2019.

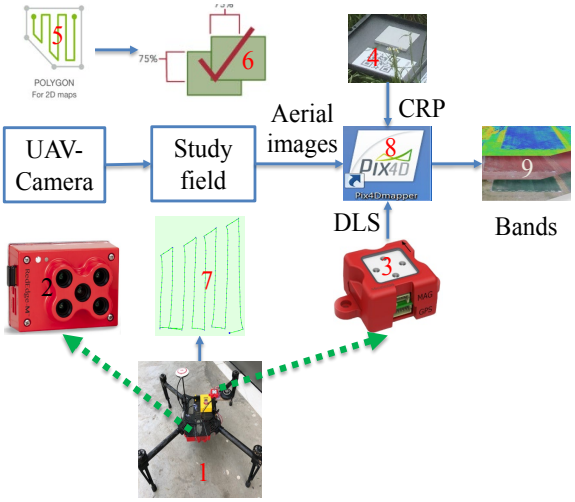

Fig. 2. UAV-Camera system for aerial imaging and image pre-processing: DJI M100 (No 1), RedEdge multispectral camera (No 2); Downwelling light sensor (No 3), reflectance calibration panel (No 4); Pix4DCapture APP for flight trajectory planning (No 5); Sidelap and overlap (No 6), flight track and imaging points (No 7); Pix4DMapper on desktop for image calibration and stitching (No 8).

ground spatial resolution is about 1.3 cm/pixel. Pix4DCapture planning software (Polygon for 2D maps) installed on a smartphone is used to plan, monitor and control the UAV. The flight track (No 7 of Fig 2), UAV forward speed (about 1 m/s) and camera triggering (see the dots in No 7 of Fig 2) are designed so that image overlap and sidelap (No 6 of Fig 2) up to 75% are achieved for an accurate orthomosaic in follow-up image processing in Pix4DMapper (No. 8 of Fig 2). Before the flight (each flight in real-life applications), reflectance calibration panel (No 4 of Fig 2) is imaged at 1m height so that an accurate reflectance data can be obtained even under environmental variations. As displayed in Fig 3, RedEdge camera equipped with GPS module can capture five narrow bands simultaneously including Blue, Green, Red, RedEdge and NIR. In addition, the necessary information for image stitching are also embedded in each image such as camera information and position/altitude information.

Upon image collection, a number of image preprocessing steps are then performed offline to produce calibrated and

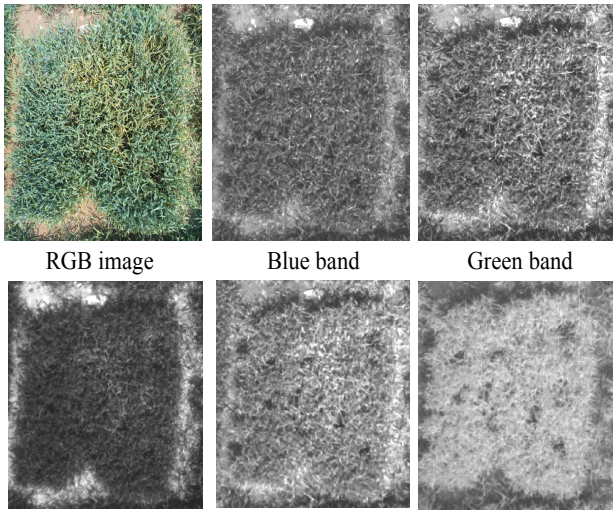

Fig. 3. Example image including RGB image taken by a small Parrot Anafi Drone at 2m above ground and five narrow bands taken by multi-spectral RedEdge camera on-board DJI M100 at 20m above ground.

georeferenced reflectance data for each spectral band. These steps are conducted in Pix4Dmapper software of version 4.3.33 (No 8 of Fig 2) including initial processing (e.g. keypoint computation for image matching), orthomosaic generation and reflectance calibration for each band [7]. The outputs are five GeoTIFF images of the covered area (No 9 of Fig 2), where the region of interest (ROI) can be cropped for follow-up analysis.

C. Data labelling

The challenge of monitoring and quantifying yellow rust disease in wheat field is formulated as a supervised multi-class classification problem. There are generally three classes in total in the region of interest, including plants with visible yellow rust lesions (Rust), healthy wheat (Healthy) and background pixels (Backg, i.e. non-vegetation soil background). In order to accurately and effectively label the multispectral image at pixel-level, a labelling flowchart is proposed, shown in Fig 4.

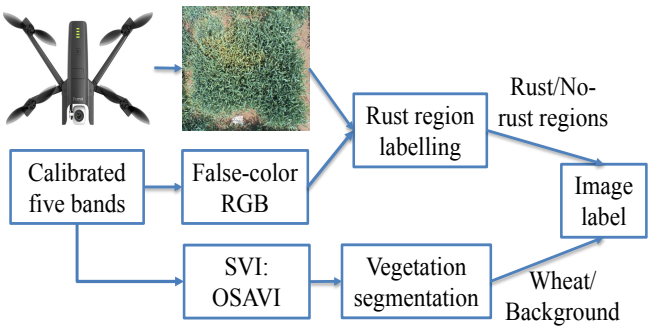

Fig. 4. Flowchart of data labelling combining rust region labelling and wheat vegetation/background segmentation.

As shown in Fig 4, the labelling of multispectral image orthomosaic relies on high-resolution proximity sensing images for visual inspection. To this end, Parrot Anafi Drone (see Fig 4) equipped with 4K HDR camera is adopted to manually take

downwards images for all 18 wheat plots. Then the steps for multispectral image labelling is summarized in Algorithm 1.

Algorithm 1: Multispectral image labelling

Input: 5 bands and 18 RGB images by Parrot Anafi.
Output: label image for each pixel.

(i) Generate false-color RGB image from the calibrated Red, Green and Blue bands; label the rust regions by Matlab ImageLabeler by inspecting the RGB images taken by Parrot Anafi, generating rust regions R_{Rust} and non-rust regions $R_{nonRust}$;

(ii) Calculate the classical Optimized Soil Adjusted Vegetation Index (OSAVI) [17] from the calibrated five bands, based on which vegetation segmentation is performed by the classical thresholding [18] to generate wheat regions R_{Wheat} and non-wheat regions $R_{nonWheat}$;

(iii) Obtain yellow rust infected wheat pixels P_{Rust}, healthy wheat pixels P_{Healty}, and background pixels P_{Backg} by the following formula

$$\begin{cases} P_{Rust} = R_{Rust} \cap R_{Wheat} \\ P_{Healty} = R_{nonRust} \cap R_{Wheat} \\ P_{Backg} = R_{nonWheat} \end{cases} . \quad (1)$$

III. Rust Monitoring System

The task of yellow rust disease monitoring in wheat field is cast as a supervised pixel-wise classification problem with three classes including Rust, Healthy and Backg. The proposed framework relies on U-Net for semantic segmentation, where the flowchart is displayed in Fig 5

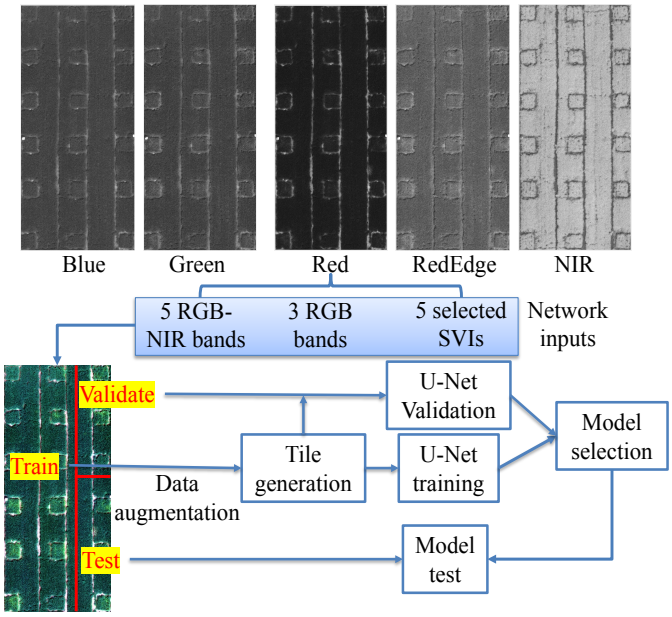

Fig. 5. Flowchart of U-Net based yellow rust semantic segmentation.

The top five images are the five calibrated bands for the RoI; the bottom left image shows the spatial split (instead of random split to test its spatial generalization) of the labelled image into training, validation and testing sub-images. Then training image tiles are generated from the training sub-image on the basis of data augmentation. Then U-Net is trained and validated to select the suitable model, which is further tested by using testing sub-image. More technical details are presented in below subsections.

A. U-Net design

The deep Convolutional Neural Network (CNN) for semantic segmentation is mainly based on U-Net, which is originally developed for biomedical image segmentation [19]. U-Net is one type of FCN [14], where no fully connected layer is used but rather based on convolution, ReLU, pooling, Up-sampling and skip connection to reduce the number of parameters for training. U-Net can take images of different sizes as its inputs and can be trained end-to-end (i.e. input: image and output: labelled image) from very few images. These characteristics are very suitable for yellow rust disease monitoring in agricultural fields due to the high cost (in terms of time and finance) in acquiring and labelling a large dataset. To make the work self-contained (in addition, the U-Net in this study is slightly different from the original one in [19]), the structure of U-Net is displayed in Fig. 6.

As shown in Fig 6, U-Net consists of a contracting path and an expansive path, where each colour block denotes a module of the network. In particular, for image input layer, zerocenter normalization (e.g. dividing each channel by its standard derivation once it has been zero-centred) is applied; each convolution process is activated by a Rectified Linear Unit (ReLU) activation function; the size of convolution kernel is 3×3 with stride [1 1] and 'same' padding; the size of max-pooling kernel is 2×2 with stride [2 2] and zero padding; the size of final convolution kernel is 1×1 with stride [1 1] and 'same' padding. The number below each block represents the size of the output image of the layer; the number above each block is the thickness of the layer. The input of the U-Net is multispectral image of image size [256, 256, No. of bands] and the output is an image with three channels representing the three classes. Considering that the number of pixels for different classes are unbalanced (making the network tend to have a low accuracy on the class with fewer samples), class-weighted cross-entropy loss function is adopted with weights inversely proportional to their frequencies

$$L = -\frac{1}{N} \sum_{n=1}^{N} \sum_{i=1}^{K} w_i T_{ni} log(Y_{ni}) \quad (2)$$

where N, K, w are number of observations, number of classes and class weight; Y, T are predicted scores and training targets.

B. Network inputs

In U-Net based semantic segmentation, various network inputs from the original five calibrated bands are compared.

Inputs A: Five RGB-NIR bands from RedEdge camera including Blue, Green, Red, RedEdge and NIR are used.

Inputs B: Three RGB bands including Blue, Green, Red are used. This is to demonstrate whether multispectral image with additional RedEdge and NIR band can provide better performance over conventional visible RGB image.

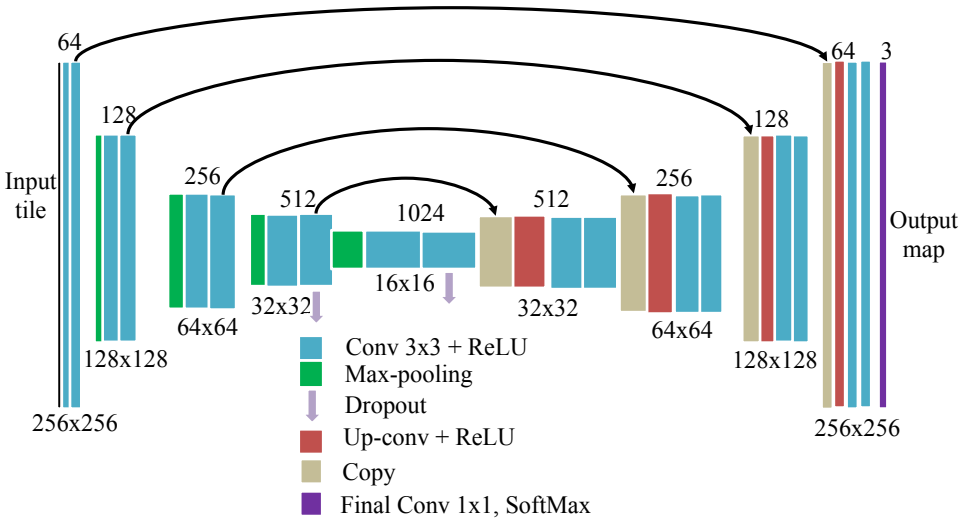

Fig. 6. The structure of the U-Net deep learning framework.

Input C: Five selected SVIs are tested. This is to show whether certain mathematical manipulations of raw bands (feature generation and selection) can further improve the performance over using raw five calibrated bands. More details (such as motivations and strategies) regarding Inputs 3 are given as below.

It has been shown in a number previous studies that some SVIs calculated from five spectral bands can provide an even higher discriminating ability in separating yellow rust diseased wheat from healthy wheat [8]. Therefore, it would be interesting to test whether better performance can be achieved by changing the network inputs from five original bands to other band combinations. To this end, in addition to the five raw calibrated bands, 18 widely used SVIs [7] are calculated and pooled with the raw five bands as the 23 candidate feature inputs. Then Sequential Forward Selection (SFS), one typical search strategy for wrappers based feature selection (see, review paper [20] for various feature selection methods), is adopted to identify the optimal band combinations (top 5 bands). In this approach, bands are sequentially added into the feature set, where the evaluation metric for adding a new band is the Out-of-Bag (OOB) error of random forest (RF) classifier. SFS with RF is summarized in Algorithm 2 [21]

Algorithm 2: SFS with RF for band selection

(a) Start with an empty set $B_0 = \emptyset, k = 0$ with full band set $B = \{b_1, \cdots, b_d\}$;

(b) Choose the next best band b^+ via

$$b^+ = \underset{b \in (B \setminus B_k)}{\arg\min}\ OOBErr(B_k \cup b),$$

where OOBErr(Y) denotes OOB error with band set Y;

(c) Update band set $B_{k+1} = B_k \cup b^+$ with $k = k + 1$;

(d) Repeat Steps (b), (c) until termination rules (i.e. objective function evaluation limit, time limit) are satisfied.

C. Network training

In this study, hundreds of multispectral images (see No 7 of Fig 2) are calibrated and stitched into one large image, which is labelled for algorithm evaluation. In particular, the image is spatially split into three sub-images for algorithm training, validation and testing. However, the training image is still too large to be segmented directly by using existing CNNs and their associated hardware. To effectively and efficiently exploit the labelled image, a large number (1600 in this work) of small image tiles of size [256 256] are randomly generated from the labelled training image, where image overlap is allowed. In image tile generation, data augmentation techniques (e.g. affine transformation from input x to output y: $y = Wx + b$) are also deployed to avoid the problem of overfitting and as to improve algorithm generalization to new scenarios [22], which include rotation within [-5 5] and scaling within [0.95 1.05] to account for various image resolutions. It is noted that online augmentation is applied on the mini-batches during training to avoid storage explosion.

The (empirically) parameter settings in optimization procedure are kept the same across different network inputs, which are summarized in Table I. The objective function is the widely-used cross entropy loss with class weighting (see, Section III-A). Regularization and drop-out are also used to tackle the problem of overfitting (i.e. improve model generalization). The hardware for network training is a GPU server equipped with an Intel(R) Xeon(R) Gold 6134 CPU@3.40GHz and an NVIDIA Tesla P100-PCIE-12GB GPU. The U-Net model is built and implemented in Matlab 2019a by using Deep Learning Toolbox, Image Processing Toolbox and Computer Vision Toolbox.

TABLE I
U-NET NETWORK PARAMETERS

Optimizer	No. Class	Momentum	Learn rate (LR)
SGDM	3	0.9	0.01
LR drop period	LR drop rate	L2Regularization	max Epochs
1	0.7	0.001	8
Mini-Batch size	Validation frequency		Tiles per epoch
16	5		1600

D. Performance metrics

To quantitatively evaluate the performance of various approaches, some widely used metrics [12] are adopted,

$$\begin{cases} Precision_c = \dfrac{TP_c}{TP_c + FP_c}, \\ Recall_c = \dfrac{TP_c}{TP_c + FN_c}, \\ F_1(c) = 2 \times \dfrac{Precision_c \times Recall_c}{Rrecision_c + Recall_c} \end{cases} \tag{3}$$

where True Positive (TP) denotes the scenario where the actual class is positive and the predicted class is also positive (i.e. correctly predicted positive values); False Positive (FP) represents the falsely predicted positive values; and False Negative (FN) is falsely predicted negative values. $Metric_c$ implies the metric value for class c. In particular, these metrics can effectively handle data with uneven class distributions. In addition to the above metric, computation time is also compared to assess the computation load where appropriate.

IV. Experimental Results

This section presents the comparative experimental results for various algorithms. Data labelling is first introduced, where the step-by-step results are shown in Fig 7. In particular, OSAVI is first applied to segment wheat pixels (white) from background pixels (black) with a threshold of 0.76 (manually determined). Then rust regions (light blue) are manually labelled in Matlab 2019a, based on which the labelled image is obtained by following the remaining steps in Algorithm 1. The labelled dataset is then applied for algorithm evaluation.

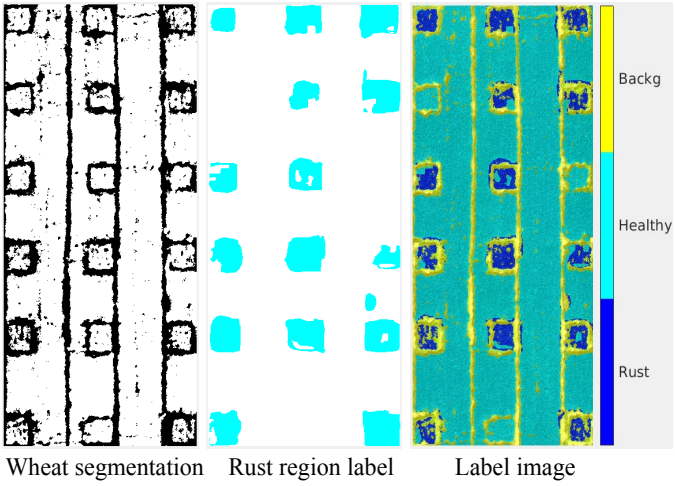

Fig. 7. Result of data labelling: wheat segmentation (left); rust region labelling (middle) and labelled image overlay on RGB image.

A. U-Net with various inputs

U-Nets with three different input band combinations (see, Section III-B) are compared. In particular, the SVIs sequentially selected by Algorithm 2 are shown in Fig 8, where the vertical axis represents the out-of-bag error of random forest classifier based on the sequentially selected bands (2000 samples for each class are randomly selected for performance calculation by random forest classifier). Then the top five SVIs are selected including OSAVI (Red-NIR), SCCCI (Red-RE-NIR), CVI (Green-Red-NIR), TGI (Green-NIR) and GI (Green-Red).

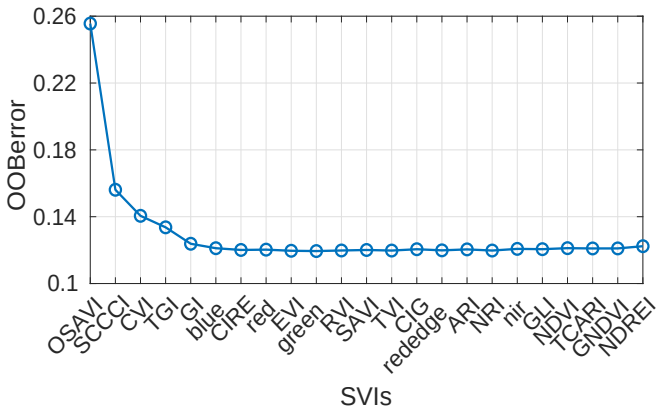

Fig. 8. SVIs sequentially selected by Algorithm 2.

The parameter setting in training is kept the same for all U-Nets with various inputs. A total of 8 epochs are adopted in each network, where each epoch contains 100 iterations. The training time using the hardware in Section III-C is about 32 minutes. Without loss of any generality, the accuracy and loss against iteration for Input A. five raw spectral bands are displayed in Fig 9. It follows from Fig 9, the accuracy increases quickly with iteration and converges after 400 iterations.

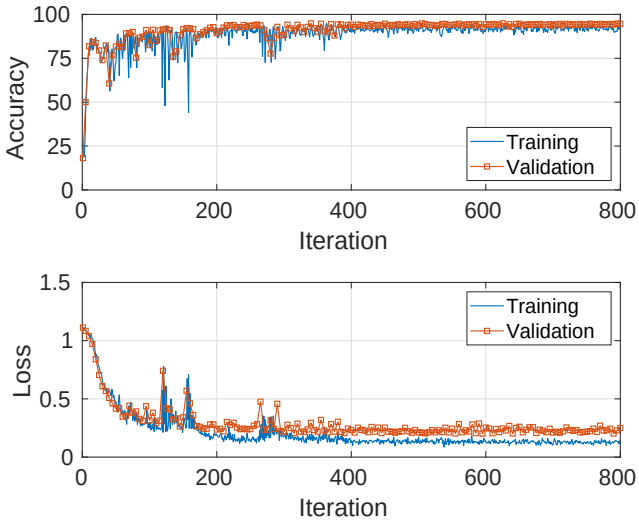

Fig. 9. Accuracy, loss against iteration for U-Net with five spectral bands.

The performance metrics for Input A, B and C are calculated on testing dataset and summarized in Tables II, III and IV.

TABLE II
Performance of U-Net with Input A.

Metric/Class	Rust	Healthy	Backg	Average
Precision	81.9%	97.9%	94.2%	91.3%
Recall	85.5%	96.3%	96.0%	92.6%
F_1 score	0.84	0.97	0.95	0.92

The trained models are also applied to training, validation and testing images with classification results in Fig 10.

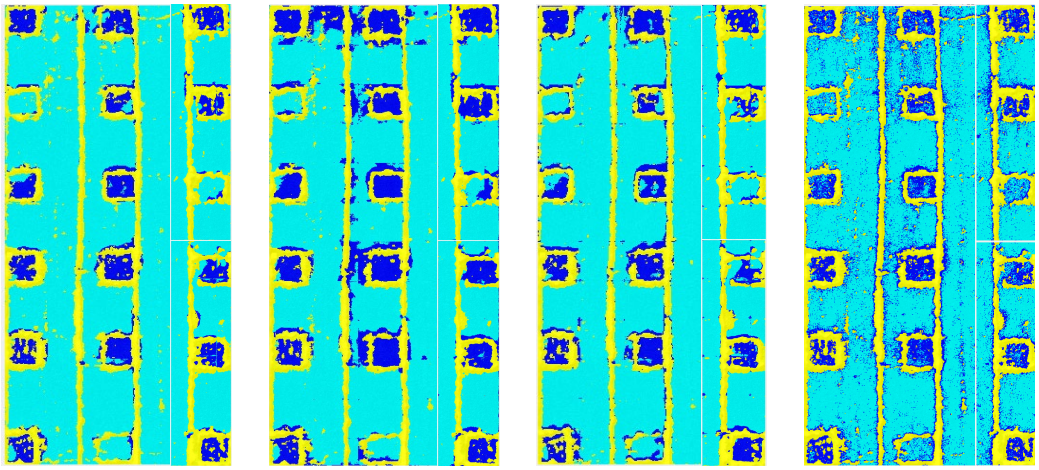

Fig. 10. Classification results on labelled image by U-Net with Input A, B, and C and spectral classifier by random forest (left to right).

TABLE III
PERFORMANCE OF U-NET WITH INPUT B.

Metric/Class	Rust	Healthy	Backg	Average
Precision	61.1%	97.8%	96.3%	85.1%
Recall	91.4%	93.2%	87.7%	90.8%
F_1 score	0.73	0.95	0.92	0.87

TABLE IV
PERFORMANCE OF U-NET WITH INPUT C.

Metric/Class	Rust	Healthy	Backg	Average
Precision	64.8%	96.1%	91.3%	84.1%
Recall	77.4%	94.3%	88.0%	86.6%
F_1 score	0.71	0.95	0.90	0.85

The following observations are drawn from above results:

(i) First comparing the results of Input A and B (with a same spatial resolution), the introduction of extra RedEdge and NIR bands can improve the classification performance. This can also be shown by the data visualization by t-SNE algorithm [23] in Fig. 11, where the data by five spectral bands obtains a better data separation for different classes.

(ii) Different from purely spectral based classification [7], selected SVIs do not improve the performance over five raw spectral bands. This is because deep learning approach can automatically learn deep features from the raw spectral bands and the selected SVIs are actually combinations of four bands including Green, Red, RedEdge and NIR.

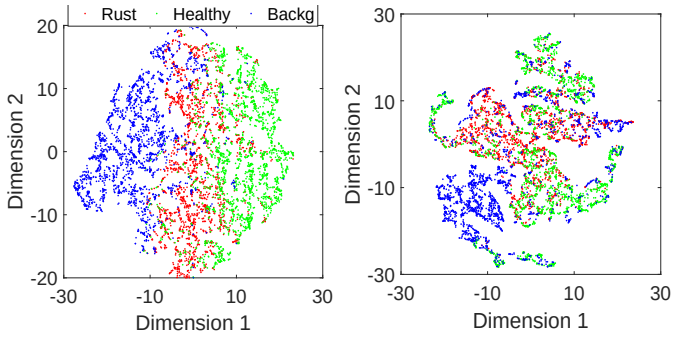

Fig. 11. t-SNE data visualization: five bands (left), RGB bands (right).

B. Spectral segmentation by random forest

The purely spectral based classifier by random forest algorithm is also tested for comparison. Considering the data imbalance problem, 20000 samples for each class are randomly selected for model training. In building the random forest classifier with tree number 100, model hyperparameters including minimum leaf size $minLS$ and number of points to split $numPTS$ are optimized by Bayesian optimization [24], where out-of-the-bag error is selected as the objective function [7]. The optimized values are $minLS = 20$ and $numPTS = 3$. Under the above parameter setting, random forest classifier is trained and its performance is evaluated on testing dataset, summarized in Table. V

TABLE V
PERFORMANCE OF RANDOM FOREST CLASSIFIER.

Metric/Class	Rust	Healthy	Backg	Average
Precision	48.9%	97.9%	97.9%	81.6%
Recall	87.4%	84.6%	98.2%	90.1%
F_1 score	0.63	0.91	0.98	0.84

Comparing the results against the ones of U-Net approaches in Section IV-A, it can be seen that U-Net approaches by automatically learning spectral-spatial features outperform purely spectral based classifier in term of F_1 score and in particular U-Net with five VIS-NIR bands excels in all metrics including Precision, Recall and F_1 score. Purely spectral based classifier also leads to a very noisy classification result.

V. CONCLUSIONS

This work aims to exploit aerial visual perception for yellow rust disease monitoring in winter wheat. An automated rust disease monitoring framework is proposed by seamlessly integrating UAV remote sensing, multispectral imaging and U-Net deep learning network. A field study is performed to generate an open-access dataset, which is applied to validate the developed framework under various network inputs and against conventional spectral based classification by random forest algorithm. Comparative results show that: (i) the introduction of RedEdge and NIR bands in multispectral image can improve segmentation performance over RGB visible image;

(ii) spectral vegetation indices do not provide better performance than raw five bands due to information loss in indices selection; (iii) U-Net deep learning based segmentation drawing spectral-spatial features concurrently outperforms purely spectral based classification by random forest. Therefore, U-Net with raw five calibrated VIS-NIR bands are preferred. Although the developed framework has been initially validated by field experiments with promising performance, there is still much room for further development, summarized below:

(i) To address data scarcity, in addition to acquiring more labelled data, data augmentation techniques should also be exploited as a more cost-effective way such as generative adversarial networks (GANs) [25, 26].

(ii) Various FCN networks and other advanced networks may also be exploited to further improve the performance such as mask R-CNN, DeepLab and their variants.

REFERENCES

[1] X. Han, H. Liu, F. Sun, and X. Zhang, "Active object detection with multi-step action prediction using deep q-network," *IEEE Transactions on Industrial Informatics*, 2019.

[2] J. Yang, K. Sim, X. Gao, W. Lu, Q. Meng, and B. Li, "A blind stereoscopic image quality evaluator with segmented stacked autoencoders considering the whole visual perception route," *IEEE Transactions on Image Processing*, vol. 28, no. 3, pp. 1314–1328, 2018.

[3] J. Yang, Y. Zhu, B. Jiang, L. Gao, L. Xiao, and Z. Zheng, "Aircraft detection in remote sensing images based on a deep residual network and super-vector coding," *Remote Sensing Letters*, vol. 9, no. 3, pp. 228–236, 2018.

[4] B. Jiang, J. Yang, Z. Lv, and H. Song, "Wearable vision assistance system based on binocular sensors for visually impaired users," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 1375–1383, 2018.

[5] M. S. Hossain, M. Al-Hammadi, and G. Muhammad, "Automatic fruit classification using deep learning for industrial applications," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 2, pp. 1027–1034, 2018.

[6] X. Zhang, L. Han, Y. Dong, Y. Shi, W. Huang, L. Han, P. González-Moreno, H. Ma, H. Ye, and T. Sobeih, "A deep learning-based approach for automated yellow rust disease detection from high-resolution hyperspectral uav images," *Remote Sensing*, vol. 11, no. 13, p. 1554, 2019.

[7] J. Su, C. Liu, M. Coombes, X. Hu, C. Wang, X. Xu, Q. Li, L. Guo, and W.-H. Chen, "Wheat yellow rust monitoring by learning from multispectral uav aerial imagery," *Computers and electronics in agriculture*, vol. 155, pp. 157–166, 2018.

[8] J. Su, C. Liu, X. Hu, X. Xu, L. Guo, and W.-H. Chen, "Spatio-temporal monitoring of wheat yellow rust using uav multispectral imagery," *Computers and electronics in agriculture*, vol. 167, p. 105035, 2019.

[9] W. Liu, G. Yang, F. Xu, H. Qiao, J. Fan, Y. Song, and Y. Zhou, "Comparisons of detection of wheat stripe rust using hyperspectral and uav aerial photography," *Acta Phytopathol. Sinica*, vol. 48, no. 2, pp. 223–227, 2018.

[10] K. Lin, L. Gong, Y. Huang, C. Liu, and J. Pan, "Deep learning-based segmentation and quantification of cucumber powdery mildew using convolutional neural network," *Frontiers in plant science*, vol. 10, p. 155, 2019.

[11] H. Huang, J. Deng, Y. Lan, A. Yang, X. Deng, and L. Zhang, "A fully convolutional network for weed mapping of unmanned aerial vehicle (uav) imagery," *PloS one*, vol. 13, no. 4, p. e0196302, 2018.

[12] I. Sa, Z. Chen, M. Popović, R. Khanna, F. Liebisch, J. Nieto, and R. Siegwart, "weednet: Dense semantic weed classification using multispectral images and mav for smart farming," *IEEE Robotics and Automation Letters*, vol. 3, no. 1, pp. 588–595, 2017.

[13] T. Zhao, Y. Yang, H. Niu, D. Wang, and Y. Chen, "Comparing u-net convolutional network with mask r-cnn in the performances of pomegranate tree canopy segmentation," in *Multispectral, Hyperspectral, and Ultraspectral Remote Sensing Technology, Techniques and Applications VII*, vol. 10780. International Society for Optics and Photonics, 2018, p. 107801J.

[14] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[15] A. Sharma, X. Liu, X. Yang, and D. Shi, "A patch-based convolutional neural network for remote sensing image classification," *Neural Networks*, vol. 95, pp. 19–28, 2017.

[16] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2961–2969.

[17] G. Rondeaux, M. Steven, and F. Baret, "Optimization of soil-adjusted vegetation indices," *Remote sensing of environment*, vol. 55, no. 2, pp. 95–107, 1996.

[18] E. Hamuda, M. Glavin, and E. Jones, "A survey of image processing techniques for plant extraction and segmentation in the field," *Computers and Electronics in Agriculture*, vol. 125, pp. 184–199, 2016.

[19] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[20] J. Tang, S. Alelyani, and H. Liu, "Feature selection for classification: A review," *Data classification: Algorithms and applications*, p. 37, 2014.

[21] P. Pudil, J. Novovičová, and J. Kittler, "Floating search methods in feature selection," *Pattern recognition letters*, vol. 15, no. 11, pp. 1119–1125, 1994.

[22] J. Wang and L. Perez, "The effectiveness of data augmentation in image classification using deep learning," *Convolutional Neural Networks Vis. Recognit*, 2017.

[23] L. v. d. Maaten and G. Hinton, "Visualizing data using t-sne," *Journal of machine learning research*, vol. 9, no. Nov, pp. 2579–2605, 2008.

[24] J. Snoek, H. Larochelle, and R. P. Adams, "Practical bayesian optimization of machine learning algorithms," in *Advances in neural information processing systems*, 2012, pp. 2951–2959.

[25] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680.

[26] M. Frid-Adar, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, "Synthetic data augmentation using gan for improved liver lesion classification," in *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*. IEEE, 2018, pp. 289–293.