

Contents lists available at ScienceDirect

International Journal of Applied Earth Observations and Geoinformation

journal homepage: www.elsevier.com/locate/jag



Detecting pine wilt disease at the pixel level from high spatial and spectral resolution UAV-borne imagery in complex forest landscapes using deep one-class classification

Jingtiao Li ^a, Xinyu Wang ^{b,*}, Hengwei Zhao ^a, Xin Hu ^a, Yanfei Zhong ^a

^a State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, PR China

^b School of Remote Sensing and Information Engineering, Wuhan University, PR China



ARTICLE INFO

Keywords:

Pine wilt disease detection
Pixel level
Deep one-class classification
Balanced unbiased detection risk
UAV hyperspectral imagery

ABSTRACT

Pine wilt disease (PWD) poses a serious threat to the worldwide pine forest resources. Unmanned aerial vehicle (UAV) remote sensing has been widely used for PWD control, due to its flexibility and efficiency. Although pixel-level detection can obtain fine detection boundaries, there have been few related works in complex scenes because of the difficulty of setting a preset category system and the poor generalization. A preset category system establishes which categories are to be labeled, and is necessary for traditional pixel-level detection. However, the poor generalization leads to an obvious accuracy drop when detecting PWD in new scenes. In the proposed approach, to address the first issue, one-class classification (OCC) is introduced to detect diseased pixels, focusing only on the category of diseased pine trees. However, the numerous objects but low PWD pixel proportion makes the model optimization unbalanced, for which balanced unbiased detection risk estimation is proposed. To address the second issue, a novel model consisting of three-dimensional (3D) convolutional layers and transformer blocks is proposed to extract more robust features. A novel PWD detection framework based on deep OCC is finally proposed to achieve fine pixel-level PWD detection results. PWD detection experiments were conducted on eight UAV H² (high spatial and spectral resolution) image strips. In total, 300 PWD samples from strip 1 (accounting for roughly 0.009 % of the total pixels) and 400 unlabeled pixels formed the training set. The test experiments were conducted in the remaining seven strips to validate the model generalization. Satisfactory quantitative results (F1-score greater than 0.9) were obtained for all the test strips. The results indicate that the proposed method has a powerful ability to detect PWD in pine trees, even when the PWD proportion is low, and shows better model generalization than the traditional pixel-level detection methods.

1. Introduction

Pine wilt disease (PWD) constitutes one of the most serious worldwide conifer diseases, which is currently affecting many countries and regions, especially in East Asia (Hao et al., 2022; Wu et al., 2020). Once the epidemic occurs, it leads to irreversible changes to forest ecosystems, increased cost of management and disease control, and restriction of international trade (Lee and Cho, 2006; Mota and Vieira, 2008; Rajasekharan et al., 2017). In recent years, increased world trade and other human activities involving the movement of wood products have increased the probability of the potential spread of PWD, and areas suitable for PWD have expanded due to global warming (Hirata et al., 2017; Ikegami and Jenkins, 2018; Matsuhashi et al., 2020). Thus,

efficient detection of PWD is both urgent and necessary. Compared to the traditional manual inspection and techniques based on satellite-borne remote sensing, unmanned aerial vehicle (UAV) remote sensing has been widely used for PWD detection in recent years because of the low cost, flexibility, and high spatial resolution (Hu et al., 2020; Lim and Do, 2021).

The majority of the PWD detection methods based on UAV remote sensing imagery are object-level methods, which locate the diseased pine trees using bounding boxes. For instance, the faster region convolutional neural network (Faster-RCNN) deep learning framework based on a region proposal network (RPN) and the ResNet neural network was used by Deng et al. (2020) to train a PWD dead tree detection model. Similar work was conducted by Hu et al. (2020), Park

* Corresponding author.

E-mail address: wangxinyu@whu.edu.cn (X. Wang).

et al. (2021), and Yu et al. (2021b). Despite the high mean average precision (mAP) that can be obtained, the rough bounding box boundaries make some of the necessary quantitative analysis impossible.

Conversely, pixel-level detection can diagnose the PWD for each pixel in the imagery without the need to consider the ratio problem, as in object-level detection. However, few previous works have explored PWD detection at the pixel level. Most of the previous studies have either achieved pixel-level results in simple scenes where only two main kinds of objects exist (i.e., healthy trees and those with PWD) (Yu et al., 2021a) or have considered complex scenes where laborious manual labeling is required for each negative object (Syifa et al., 2020). Two main difficulties can exist when detecting PWD in complex scenes at the pixel level: 1) **Laborious manual labeling is required for the numerous objects in complex forest landscapes.** For example, the pixel-level detection work of Syifa et al. (2020) required the manual labeling of seven kinds of objects, including PWD, shadows, roads, etc. This complex object distribution is common in forest landscapes, where laborious manual labeling is typically needed. Furthermore, the severe class imbalance can result in the model optimization ignoring the low-proportion PWD objects. 2) **A visible drop in accuracy can occur when detecting PWD in images that are different from the training images.** The spectral heterogeneity between the training and test images is mostly obvious due to the influence of the imaging environment and instrument error, making the knowledge learned in the training images not applicable to the test images. Furthermore, some new objects may appear in the test images, causing confusion for the trained model. The above two problems can severely limit the performance of pixel-level PWD detection in complex scenes.

To solve the first problem, one-class classification (OCC) (Bellinger et al., 2012; Perera et al., 2021) is introduced in the proposed method, which only needs a small number of labeled PWD samples. Because OCC only focuses on the PWD and does not need the labels of the non-PWD objects, a preset category system is not necessary. However, scenes with numerous objects but a low proportion of PWD are the most common in forest landscapes. This imbalance in the object distribution makes the traditional risk estimators (Du Plessis et al., 2015, 2014; Kiryo et al., 2017; Wang et al., 2016) ignore the PWD and focus on the dominant objects. To address this issue, a balanced unbiased risk estimator is proposed by copying the PWD samples to make the proportion

balanced while keeping the unbiased characteristic.

To solve the second problem, a novel model consisting of three-dimensional (3D) convolutional layers (Mäyrä et al., 2021) and transformer blocks (Dosovitskiy et al., 2020) is introduced to extract robust features, to counter the spectral heterogeneity. Specifically, the spatial and spectral features are jointly extracted with movement along both the spatial and spectral dimensions, and an attention mechanism is applied to make the model focus only on the discriminative regions.

In this paper, a novel PWD detection framework is proposed, based on deep OCC, with the aim being to achieve highly accurate pixel-level PWD detection results. A balanced unbiased risk estimator is proposed to make the optimization balanced under the low PWD proportion condition, and a deep detection model is constructed to extract robust features from the hyperspectral remote sensing images and enhance the model generalizability. PWD detection experiments were conducted on eight high spatial and spectral resolution UAV image strips, which we refer to here as H² imagery (Hu et al., 2022; Zhong et al., 2020). To the best of our knowledge, this study is the first to apply deep OCC to forest disease detection.

2. Materials

2.1. Study area and data acquisition

The study area for the PWD detection was Yantai Hill, Hexi Village, Shandong province, China (see Fig. 1). The longitude of the study area ranges from 121.884° to 121.889° and the latitude ranges from 37.327° to 37.331°, covering 0.312 km². The altitude is 125.21 m, the annual rainfall is 750 mm, and the average temperature is 12 °C. The main pine species in the study area are *Pinus massoniana* Lamb, *Pinus armandii* Franch, and *Pinus parviflora*.

Data acquisition was achieved using a Headwall Nano-Hyperspec imager mounted on a DJI M600 Pro UAV. The hyperspectral sensor can obtain 274 bands, in practice, with an imaging range of 400–1000 nm and a spatial resolution of 11 cm. The acquisition time was from 12:30 PM to 14:00 PM on September 17, 2020, when eight UAV H² image strips were captured. Although the sizes of the eight strips were slightly different, they were all approximately 4600 × 700 pixels, with about 30 % overlap between adjacent strips. The individual strips are

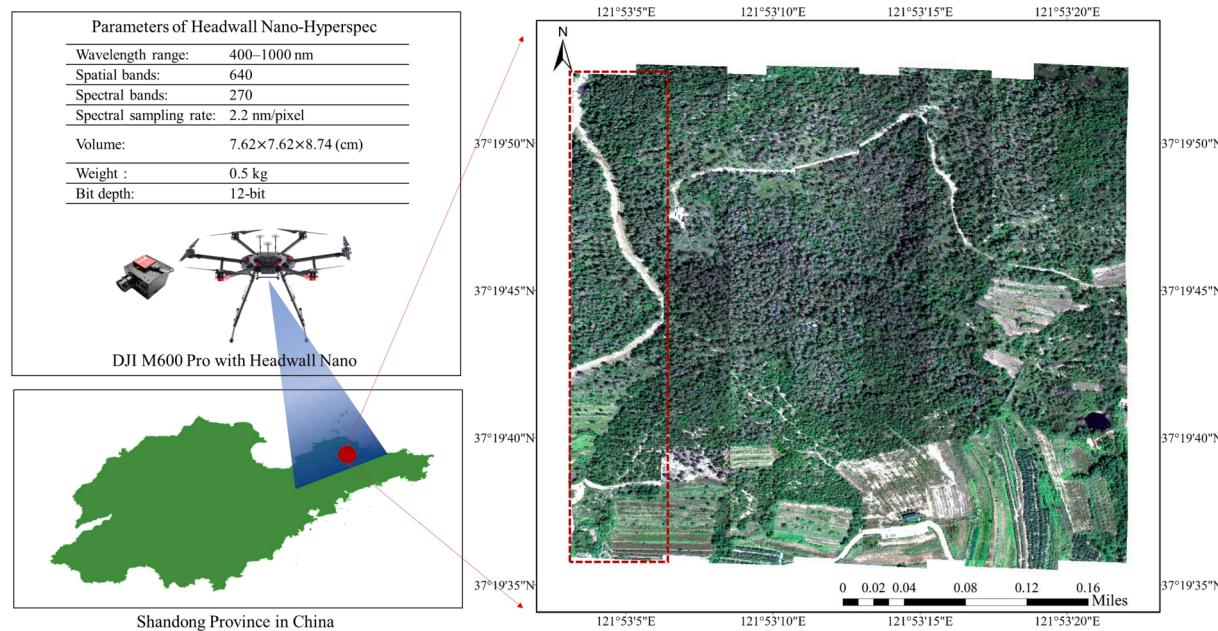


Fig. 1. The study area, which covers 311,696 m², is located in Hexi Village, Shandong province, China. The study area was covered by eight H² UAV strips. The highlighted strip (with the red border) was used to make the training set, and the other strips formed the test set.

visually distinct in the mosaicked image, and are named 1–8, from left to right, for convenience.

2.2. Challenges for pixel-level PWD detection

The detection task in this study was focused on the middle stage (yellow needles) and late stage (reddish-brown needles) (Yu et al., 2021b) of PWD. Compared with general detection scenes, there are two main challenges that would be encountered in this study area when detecting PWD using the traditional pixel-level detection methods: 1) The numerous objects make the category system difficult to set, and the object distribution is highly imbalanced. The objects in the study area include healthy pine trees, diseased pine trees, roads, bare land, houses, lake, grassland, bushes, soil, etc. The large number of objects represents a challenge for setting the category system and labeling the negative samples (i.e., the objects except for PWD). Moreover, the object distribution is highly imbalanced. For the diseased pixels, the statistical proportion in each strip is shown in Table 1, where three of the proportions are less than 0.1. The severely imbalanced distribution leads to the detection model optimization favoring fitting the negative samples. 2) The eight strips were not captured at the same time period, and the object kinds vary from strip to strip, making it difficult to detect PWD in image strips that are different from the training image strip. The spectral heterogeneity and the unseen objects in the test strips make the knowledge learned in the training strip not applicable, limiting the generalizability.

3. Methods

The proposed framework consists of three basic components, as shown in Fig. 2. To avoid the problem of setting a category system and to strengthen the PWD detection generalizability, OCC is adopted as the basic principle of the proposed framework, which focuses on only the PWD samples (Bellinger et al., 2012; Piironen et al., 2018; Sabokrou et al., 2020). To deal with the imbalance problem when using OCC in PWD detection, as discussed in Section 2.2, the balanced unbiased detection risk estimator is proposed to make the optimization balanced, even in a low PWD condition. In addition, a deep learning technique is used to construct the detection model, with 3D convolutional layers and transformer blocks, to enhance the detection ability.

In the proposed framework, PWD proportion estimation is the first component, which outputs the estimated initial PWD proportion to the second component, i.e., the balanced unbiased PWD detection risk estimation. The third component is the detection model, which extracts the deep hyperspectral spatial and spectral features to judge the diseased pixels under the balanced unbiased PWD detection risk estimation.

3.1. PWD proportion estimation

According to the research of De Comité et al. (1999), the background objects in UAV imagery have the potential to improve the detection accuracy by exploiting them to optimize a model jointly with the PWD samples in OCC. However, the PWD proportion in the imagery needs to be known in advance (Denis et al., 2005), which is elaborated in the next section. Inspired by this, kernel mixture proportion estimation (KMPE) (Ramaswamy et al., 2016) was utilized to estimate the PWD proportion in this study. As shown in Fig. 2, KMPE has two inputs: the PWD sample pixels and the randomly selected pixels of the background mixture objects. KMPE maps these two obtained distributions of PWD and background mixture to a reproducing kernel Hilbert space (RKHS), aiming to

find the most suitable ratio that can reconstruct the actual mixture distribution. This component outputs the estimated PWD proportion for the construction of the other component, i.e., the balanced unbiased PWD detection risk estimation.

3.2. The balanced unbiased risk estimator for PWD detection

We let $X^d \in R^d$ be the training data, containing n_p labeled positive samples (PWD) and n_u unlabeled samples (including both PWD and non-PWD samples). π_n and π_p are the proportions of the negative samples (non-PWD) and positive samples in the unlabeled samples, respectively, where $\pi_n = 1 - \pi_p$. Once π_p is known, the number of PWD and non-PWD samples in the unlabeled samples can be estimated to be $n_u\pi_p$ and $n_u\pi_n$. $g : R^d \rightarrow R$ represents the decision function mapping input data to the probability of being PWD, and l is the loss function of the detection model. $\hat{R}_p^+(g) = (1/n_p) \sum_{i=1}^{n_p} l(g(x_i), 1)$ measures the risk of classifying the positive samples as PWD. $\hat{R}_u^-(g) = (1/n_u) \sum_{i=1}^{n_u} l(g(x_i), 0)$ represents the risk of classifying all the unlabeled samples as non-PWD. The estimated risk for the n_p positive samples and $n_u\pi_n$ negative samples in the unlabeled samples is denoted as $\hat{R}_{pn}(g)$. The traditional unbiased risk estimation (Du Plessis et al., 2014) is shown in Eq. (1):

$$\hat{R}_{pn}(g) = \pi_p \hat{R}_p^+(g) + \hat{R}_u^-(g) - \pi_p \hat{R}_p^-(g) \quad (1)$$

However, if the PWD proportion value is small (i.e., having a gap in orders of magnitude compared to the non-PWD objects), the estimated positive risk of PWD would be negligible, compared to the negative risk of other non-PWD objects. Thus, the model optimization process would be severely biased toward the negative objects, rather than PWD. To solve this imbalance, the balanced unbiased risk estimator is proposed. The mathematical expression for this is as follows:

$$\hat{R}_{pn}(g) = \frac{\hat{R}_p^+(g)}{2} + \frac{\hat{R}_u^-(g) - \pi_p \hat{R}_p^-(g)}{2(1 - \pi_p)} \quad (2)$$

The hypothesis is that, if the PWD proportion value is small or large (i.e., an order of magnitude difference), it can be adjusted manually by copying or deleting PWD samples. According to the work of Du Plessis et al. (2014), $\hat{R}_u^-(g) - \pi_p \hat{R}_p^-(g)$ is actually equal to $\pi_n \hat{R}_n^-(g)$, which is the risk of classifying the negative samples in the unlabeled samples as non-PWD. $\hat{R}_p^+(g)$ and $\hat{R}_n^-(g)$ have a ratio of $\pi_p : \pi_n$ in Eq. (1). To proportionally control this to 1 : 1, the estimated risk of the PWD samples needs to be multiplied by an adjustment factor $\frac{1-\pi_p}{\pi_p}$. Fig. 3 shows the intuitive principle of balanced unbiased risk estimation when adjusting the risk ratio by copying the PWD samples. From the unlabeled area, the ratio of the PWD samples to other object samples can be estimated, which is 1:2 in Fig. 3. Then, based on the estimated proportion, the PWD samples in the labeled PWD area are copied to make the ratio with the negative samples in the unlabeled samples 1:1. The adjusted risk estimation can be seen as a special case of Eq. (1), which guarantees the unbiased property of the proposed balanced risk.

Inspired by the work of Kiryo et al. (2017), to prevent the overfitting problem when the negative risk estimation is less than 0, the negative risk in Eq. (3) is corrected to a non-negative risk, and the final balanced unbiased risk estimator is obtained as follows:

Table 1
True proportion of diseased pixels in the eight strips.

	1	2	3	4	5	6	7	8
Proportion	0.041	0.140	0.148	0.114	0.053	0.057	0.145	0.170

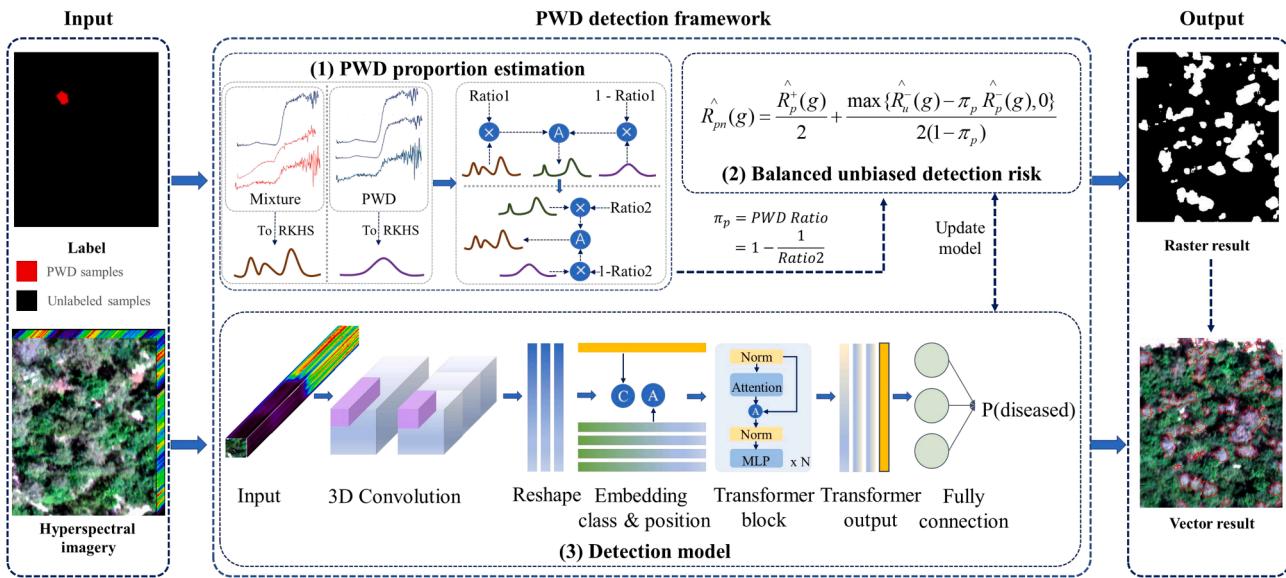


Fig. 2. The proposed PWD detection framework, which consists of three main components: (1) PWD proportion estimation; (2) balanced unbiased detection risk estimation; and (3) the detection model. Firstly, the PWD proportion in the strip is estimated using the labeled PWD samples and the randomly selected mixture samples (including both PWD and non-PWD objects). The balanced unbiased detection risk is then estimated and acts as the criterion to optimize the model. Note that only a small number of PWD samples need to be labeled.

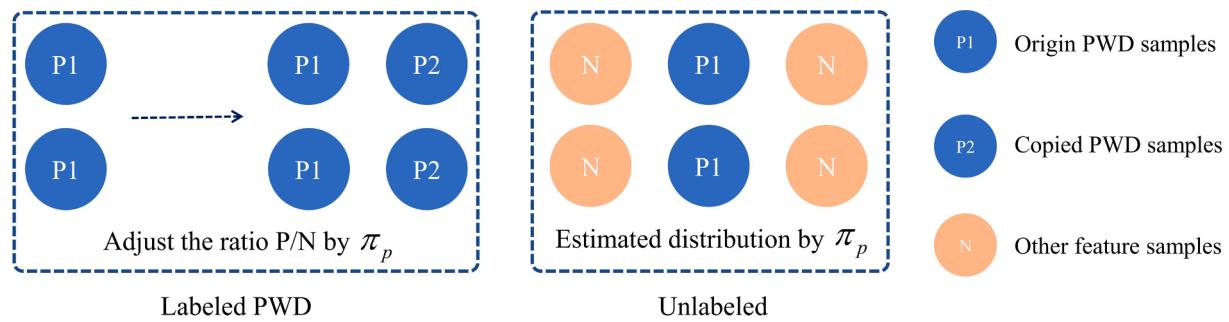


Fig. 3. The basic idea of the proposed balanced unbiased detection risk. The balance is achieved by adjusting the proportion of the PWD samples and the negative samples in the unlabeled samples to be 1:1 using the estimated π_p .

$$\hat{R}_{pn}^+(g) = \frac{\hat{R}_p^+(g)}{2} + \frac{\max\{\hat{R}_u^-(g) - \pi_p \hat{R}_p^-(g), 0\}}{2(1 - \pi_p)} \quad (3)$$

3.3. Detection model architecture

The proposed detection model is composed of two main parts—3D

convolutional layers and transformer blocks (Dosovitskiy et al. 2020)—which are shown in Fig. 2. Compared to a two-dimensional (2D) convolutional layer, as used in many hyperspectral applications Liu et al., 2021a, 2021b; Yang et al., 2018), a 3D convolutional layer is better able to extract the spatial-spectral features of the hyperspectral data cube (Mäyrä et al., 2021). The reason for introducing the transformer blocks is to encourage the model to pay more attention to the critical features

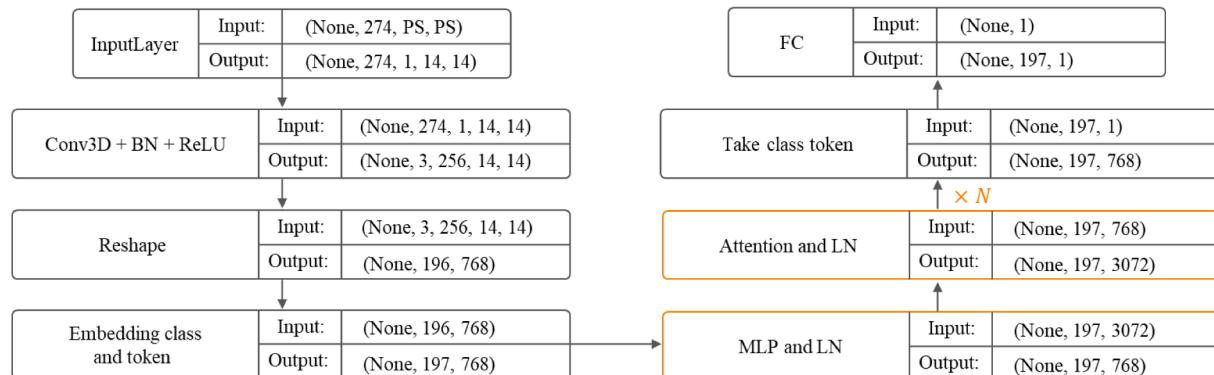


Fig. 4. Tensor change in the detection model corresponding to Fig. 3. PS: patch size; BN: batch normalization; ReLU: rectified linear unit; MLP: multilayer perceptron; LN: layer normalization; FC: fully connected.

and to correlate the global features. The detailed change flow of the inner tensor is given in Fig. 4. The patch size (PS) is a hyper-parameter to control the spatial size of the input patch cube. The 3D convolutional layers first act as the feature extractor for the input with size $PS \times PS \times 274$ in the first layer (Li et al., 2017), and then the extracted features are reshaped to a 2D representation to feed into the transformer block. After being processed in the transformer block for two iterations, the class token is taken from the transformer output and followed by a fully connected layer, to give the final PWD detection probability. The detection model was implemented using PyTorch 1.10.1 in this study (Imambi et al., 2021).

3.4. Experimental settings

3.4.1. Sample selection

To guarantee the geographic isolation characteristic, as in the real world, strip 1 was used to make the training set, and the remaining strips formed the test set, as shown in Fig. 1. To demonstrate the advantage of the proposed framework in reducing the labeling workload, only 300 diseased pixels (as discussed in Section 4.3.2) in strip 1 were labeled (accounting for roughly 0.009 % of the total pixels) and 400 unlabeled pixels were selected randomly to form the training set. The test set contained 10,000 PWD pixels and 10,000 other object pixels in each test strip. For each pixel, its surrounding spatial information was incorporated to form a patch with the size of 15×15 pixels (as discussed in Section 4.3.1) as the input.

3.4.2. Comparison methods

The OCC detection models used for the comparison were one-class support vector machine (OCSVM) (Schölkopf et al., 1999), one-class biased support vector machine (BSVM) (Piiroinen et al., 2018), and deep support vector data description (Deep SVDD) (Ruff et al., 2018). OCSVM and BSVM are classical OCC models based on the support vector machine (SVM) mechanism. As an important breakthrough, Deep SVDD has successfully introduced a deep learning model into the OCC community using only labeled positive samples. Although some recent deep OCC models have achieved satisfactory results using both labeled and unlabeled samples (Lei et al., 2021; Zhao et al., 2022), the used risk estimators have little ability to deal with the imbalance problem, and fail in the PWD detection task. Thus, these models were not compared in the experiments conducted in this study.

3.4.3. Evaluation metrics

Three PWD evaluation metrics are considered in this paper—precision (P), recall (R), and F1-score—for which the formula expressions are given below. The true-positive (TP), false-positive (FP) and false-negative (FN) pixels are the basic elements used to compute the metrics. The precision reflects the correct detection rate in pixels that are certified as PWD, and the recall shows how many PWD pixels have been correctly detected. Because the precision and recall are always changing in opposite directions, the F1-score is introduced to measure the comprehensive detection capability for PWD and other objects.

$$P = \frac{TP}{TP + FP} \quad (4)$$

$$R = \frac{TP}{TP + FN} \quad (5)$$

$$F1 = \frac{2PR}{P + R} \quad (6)$$

4. Results

4.1. Quantitative analysis of the detection results

The proportion of diseased pixels in the training strip (i.e., strip 1)

was first estimated, using the positive part in the training set as the PWD sample input, and 2000 pixels were randomly selected from the strip as the background mixture input. The obtained estimated proportion was 0.013, which represents an order of magnitude difference in the ratio of PWD to non-PWD, and shows a slight deviation from the true proportion listed in Table 1. The low and inaccurate estimation brings great challenges to the following PWD detection. Then, based on the estimated proportion, the balanced unbiased risk was estimated, and the proposed framework was applied for PWD detection in each test strip.

The quantitative evaluation results are listed in Table 2, in which the red, green, and blue figures represent the best precision, recall, and F1-score values in each strip, respectively. Compared to OCSVM, BSVM obtains better precision and recall scores, which shows the benefit of using unlabeled objects to improve the PWD detection performance. Nevertheless, BSVM is heavily dependent on a well-designed hyper-parameter selection process, limiting its practical application. Deep SVDD shows a significant improvement in all three metrics, compared to BSVM. This improvement implies that the matrix representation of the spatial information is superior to the vector form, and the deep features have a better discrimination ability and generalization power than linear features. The proposed framework achieves the highest recall and F1-score values in all seven strips, which are all above 90 %. As the recall rate is much more important than the precision in practical detection, it can be considered that the proposed framework is suitable for use in real applications.

Another great advantage of the proposed framework is that no threshold needs to be set, compared to BSVM and Deep SVDD, because 0.5 is a natural threshold, due to the use of the sigmoid activation function in the last layer. In conclusion, the quantitative comparison shows that the proposed framework is superior to the existing one-class detection models when dealing with the PWD detection task.

4.2. Qualitative analysis of the detection results

The qualitative results are shown in Fig. 5, where the white region represents the diseased pine trees. Consistent with the high-precision and low-recall results in Table 2, OCSVM identifies the fewest diseased pixels in the red rectangle area. Conversely, more correct diseased regions are discriminated by BSVM, which leads to both the high precision and high recall in Table 2. Due to the limited image feature representation ability of a one-dimensional vector, the detection results of both OCSVM and BSVM have some noisy points and are not sufficiently smooth. Deep SVDD identifies most regions as diseased pines, and most of the non-diseased pixels in the red rectangle area are misclassified. From this comprehensive comparison, the proposed framework shows a better diagnostic capability and balance, in view of the smoothness and accuracy.

Moreover, the qualitative detection results of the proposed detection framework were converted to polygon vector format, and are displayed in Fig. 6 (partially enlarged) and Fig. 7 (large scale). From the perspective of both the partially enlarged and large-scale visual effects, although only a few diseased pixels were labeled, without labeling the various negative objects, the proposed framework can still distinguish the diseased pine trees well from the many other objects (i.e., road, grass, bare areas, healthy trees, etc.) and can obtain fine detection boundaries.

4.3. Sensitivity analysis

4.3.1. Detection accuracy sensitivity to patch size (area)

As is well known, patches are always adopted as the input mode when completing pixel-level discrimination, so as to use the rich spatial features. A larger patch size means more spatial information to use (Cao et al., 2018; Zhang et al., 2016), but also increased computational complexity, interference, and training time. For a given UAV image, the appropriate patch size is related to its spatial resolution, spectral

Table 2

Quantitative comparison of the detection results for the test set made up of seven image strips, where the red, green, and blue figures represent the best precision, recall, and F1-score in each strip. The proposed framework achieved the highest recall and F1-score values in all the test strips.

Strip	OCSVM			BSVM			Deep SVDD			Ours		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
2	1.000	0.302	0.464	1.000	0.573	0.729	0.772	0.853	0.811	0.870	0.999	0.931
3	0.997	0.339	0.506	1.000	0.652	0.789	0.823	0.872	0.847	0.884	0.996	0.936
4	0.939	0.306	0.461	0.956	0.837	0.893	0.821	0.905	0.861	0.998	0.932	0.964
5	1.000	0.614	0.761	1.000	0.538	0.700	0.899	0.869	0.884	0.840	0.998	0.912
6	0.999	0.606	0.754	0.998	0.760	0.863	0.799	0.908	0.850	0.971	1.000	0.985
7	0.974	0.217	0.355	0.989	0.636	0.774	0.672	0.856	0.753	0.952	0.883	0.916
8	0.938	0.309	0.465	0.945	0.614	0.744	0.672	0.848	0.750	0.929	0.921	0.925

Strip	OCSVM			BSVM			Deep SVDD			Ours		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
2	1.000	0.302	0.464	1.000	0.573	0.729	0.772	0.853	0.811	0.870	0.999	0.931
3	0.997	0.339	0.506	1.000	0.652	0.789	0.823	0.872	0.847	0.884	0.996	0.936
4	0.939	0.306	0.461	0.956	0.837	0.893	0.821	0.905	0.861	0.998	0.932	0.964
5	1.000	0.614	0.761	1.000	0.538	0.700	0.899	0.869	0.884	0.840	0.998	0.912
6	0.999	0.606	0.754	0.998	0.760	0.863	0.799	0.908	0.850	0.971	1.000	0.985
7	0.974	0.217	0.355	0.989	0.636	0.774	0.672	0.856	0.753	0.952	0.883	0.916
8	0.938	0.309	0.465	0.945	0.614	0.744	0.672	0.848	0.750	0.929	0.921	0.925

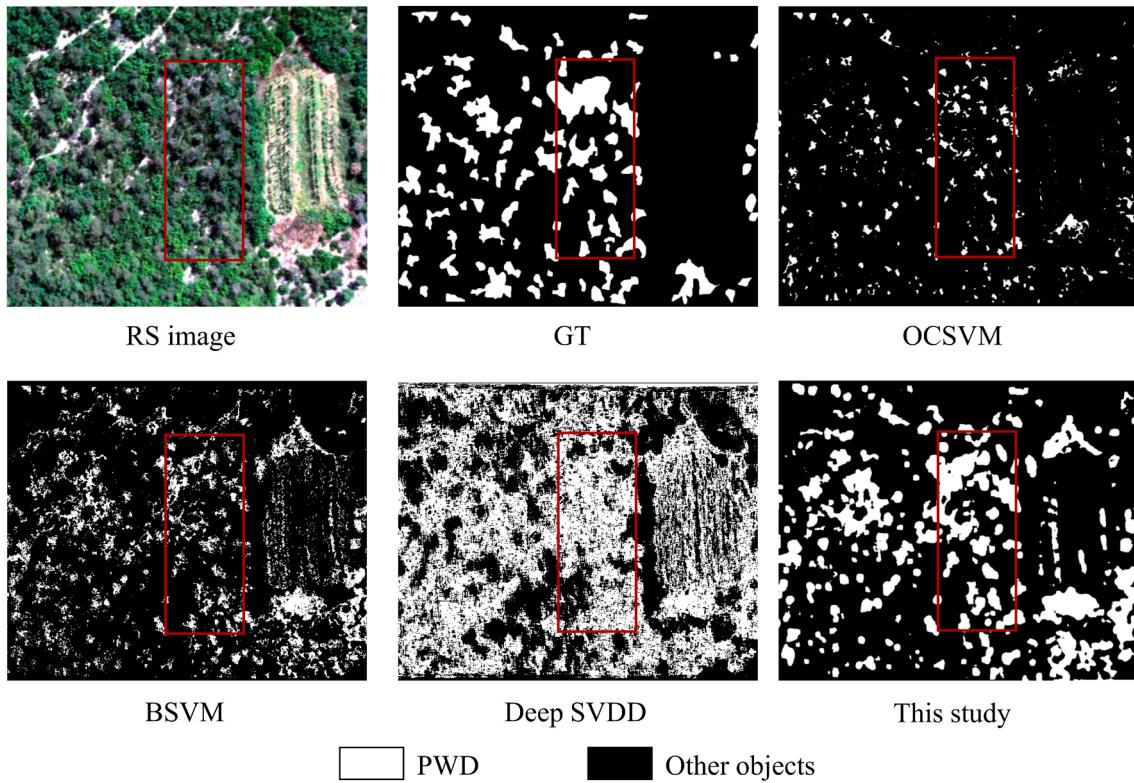


Fig. 5. Qualitative results, where the white region represents the diseased pine trees. The proposed framework achieved the closest detection map to the ground truth (GT), with less noise and a higher accuracy than the other compared models.

resolution, and feature distribution characteristics. To achieve efficient detection, choosing a suitable patch size acts as the important first step. In the study area, different patch sizes were set, i.e., 7×7 , 11×11 , $15 \times$

15 , and 19×19 , corresponding to areas of 0.59 m^2 , 1.46 m^2 , 2.72 m^2 , and 4.36 m^2 , respectively.

The results are shown in Fig. 8a. The F1-score is used to

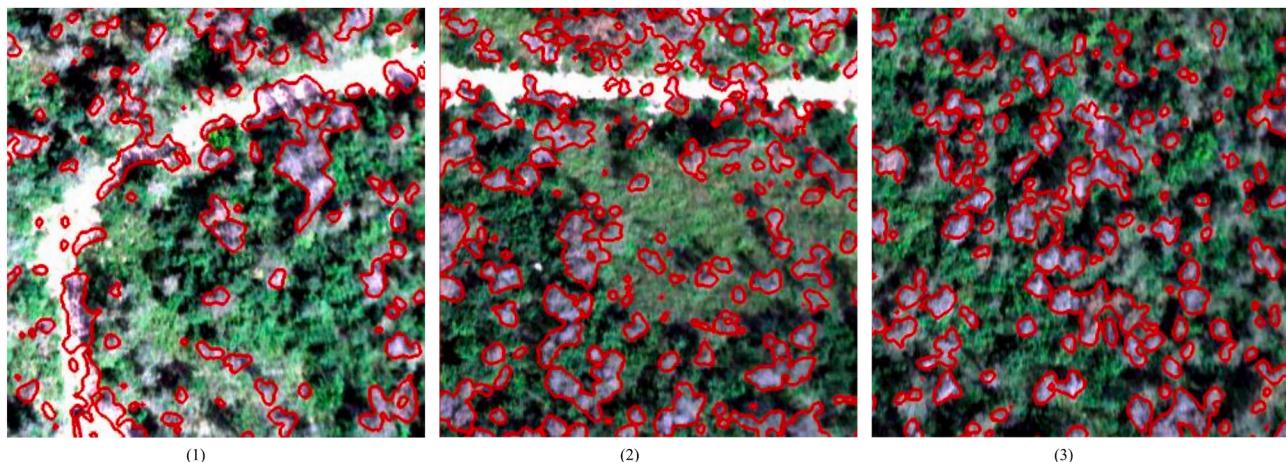


Fig. 6. Partially enlarged visual detection results, which show that fine pixel-level detection results can be obtained. The obtained diseased boundaries are smooth and accurate. Although no labeled negative samples (non-PWD objects) were given during the training, the various non-PWD objects (e.g., roads, bushes, grasses, and healthy trees) in the above three scenarios have been correctly recognized.

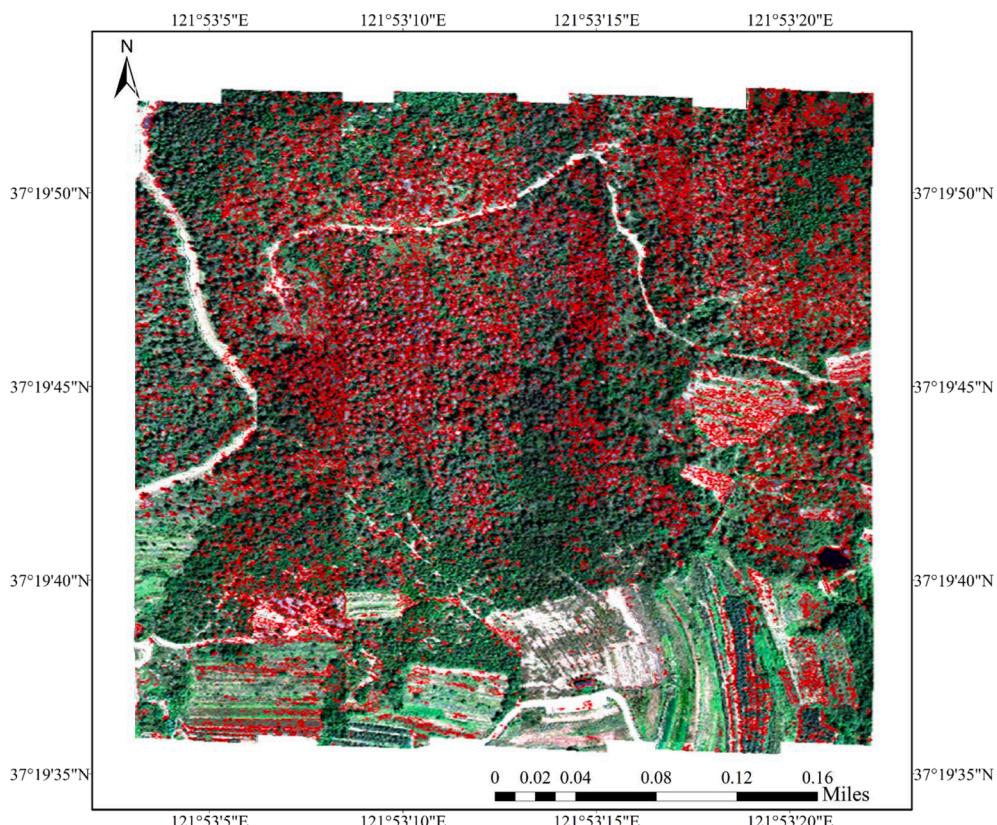


Fig. 7. Large-scale qualitative detection results of the proposed framework on the eight strips, for which the raster results have been converted to polygon vector format (the red lines). Although only 300 PWD pixels were labeled, the proposed framework can still discriminate PWD from most of the other objects.

comprehensively consider the precision and recall rate. A higher F1-score indicates a better performance in one test strip, while a tight grouping of scores between test strips indicates a more stable performance. As can be seen, higher and more stable results are obtained as the patch size increases, but this increase slows down around a patch size setting of 15×15 . The F1-score distribution is not even for patch sizes of 7×7 and 11×11 , where the range is larger than 0.2. Conversely, a tighter grouping of results is obtained for patch sizes of 15×15 and 19×19 , and almost all the strips show an F1-score of above 0.9 using a patch size of 15×15 . There is a slight drop in F1-scores between the patch sizes of 15×15 and 19×19 , which may suggest that a patch size

of 19×19 brings some interference and holds back the detection ability. Thus, the patch size of 15×15 , covering 2.72 m^2 , was chosen as the best patch size setting.

4.3.2. Detection accuracy sensitivity to labeled PWD sample number

Due to the rich spectral information of hyperspectral pixels, and the characteristic of OCC, in that it only requires labeled PWD samples (Sabokrou et al., 2020; Zhao et al., 2022), rather than a complex category system, the proposed framework has a low requirement for labeled PWD pixels. To further make this requirement clear, a comparison experiment with different numbers of labeled PWD pixels was

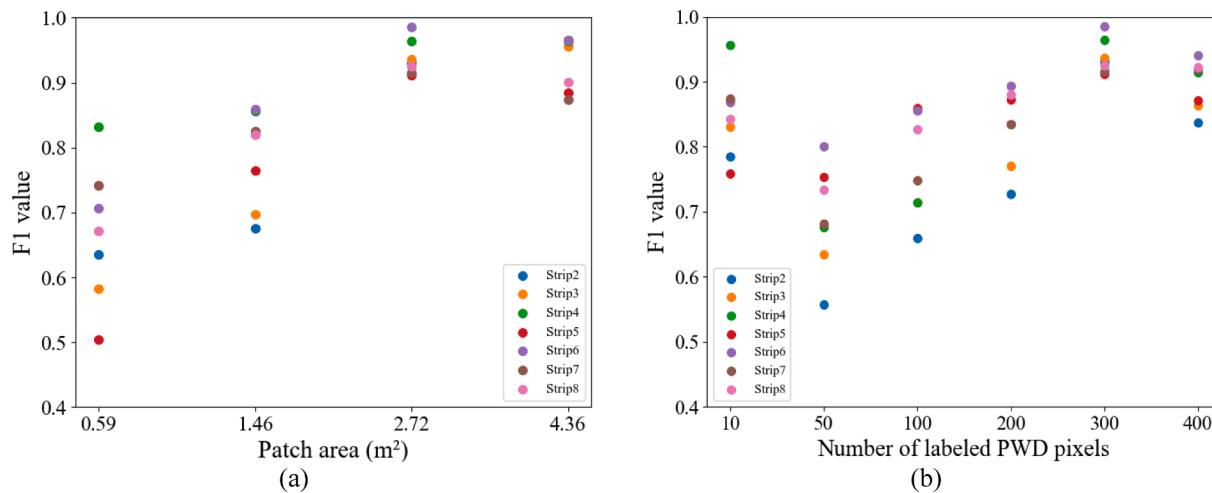


Fig. 8. Sensitivity analysis for the input patch size and the number of labeled PWD pixels for training. (a) The effect of different patch sizes on the detection F1-score. (b) The effect of different PWD sample numbers on the detection F1-score. The results show that robust F1-scores can be obtained under a patch size of 15×15 (which is equal to an area of 2.72 m^2) and a labeled pixel number of 300.

conducted, and the results are shown in Fig. 8b.

In Fig. 8b, a higher dot means a better detection performance, and tightly grouped dots mean stable and robust results obtained in the corresponding labeled PWD sample number setting. The overall trend is that a better and more robust performance can be achieved as the number of labeled PWD samples increases. In fact, the somewhat unusually good results obtained using just 10 labeled pixels also makes sense because the balanced risk estimation averages the risk in all the labeled pixels, so that the sample quality is much more important than the quantity in the proposed framework. The performance gain with an increase in quantity actually comes from the increase in overall sample quality. With 300 labeled PWD pixels, the F1-scores for the eight strips are all over 0.9, and a slightly lower detection performance is shown with 400 labeled PWD pixels. Thus, the setting of only labeling 300 PWD pixels for the proposed framework in this study area was chosen. By labeling the PWD pixels using the ROI tool in ENVI (Exelis, 2015), the labeling of the 300 samples could be completed in several seconds. Hence, the proposed framework can achieve a satisfactory performance with a low human labor cost.

4.4. PWD detection generalizability: OCC vs Traditional pixel-level detection

To show the better generalization ability of the proposed framework, 100 pixels for each kind of negative object were labeled (including healthy pine trees, road, bare soil, grassland). In addition, the balanced risk estimation was changed to cross-entropy loss (Janocha and Czarnecki, 2017), based on the current framework. The results are listed in Table 3. A high F1-score of 0.926 can be achieved in strip 1 when using traditional pixel-level detection, but the F1-scores of test strips 2–7 show an obvious drop, compared to the results in Table 2 when using traditional pixel-level detection (Haut et al., 2018). The reason for this may be the focus difference between OCC and binary classification, in that OCC only focuses on the PWD objects, but binary classification also focuses on the negative objects, in addition to PWD. Furthermore, new

negative objects may occur in other image strips and result in confusion for the traditional pixel-level detection model (Bendale and Boult, 2016). Based on the above points, it is reasonable to say that OCC has a better generalizability than general binary classification.

5. Discussion

Currently, object-detection methods are used often for PWD detection due to their simplicity and the fact that they do not require any adaptation (Deng et al., 2020; Park et al., 2021; Wu et al., 2021; Yu et al., 2021b). In contrast, pixel-level detection methods always perform poorly in complex scenes due to the category system problem and the poor generalizability, as mentioned in the Introduction section. However, PWD detection at the pixel level is necessary for the subsequent statistical analysis. To solve this dilemma was the main motivation for this study.

OCC has great potential to deal with the above problem, but has received little attention from forestry researchers, due to the differences in the research fields. Compared to the traditional pixel-level detection methods (Syifa et al., 2020), OCC only needs a small number of labeled PWD samples. The labeling workload is thus reduced, and there is no need for a category system, which is a great benefit for practical detection. Furthermore, the boundaries constructed by OCC closely surround the diseased samples, rather than a boundary between PWD samples and negative samples, leading to better model generalization, as shown in Section 4.4.

The main contribution of this study is that we introduced OCC and solved the imbalance problem when using OCC for pixel-level detection, to achieve fine pixel-level detection results. Specifically, the imbalance problem of the traditional unbiased detection risk, as mentioned in Section 3.2, makes the model optimization favor the dominant objects and results in low recall under the low PWD proportion condition, as shown in Table 1. This problem was solved by adding balance to the original detection risk by adjusting the PWD proportion while keeping the unbiased characteristic.

Table 3

The PWD detection results obtained using traditional pixel-level detection (i.e., binary classification). A high F1-score of 0.926 is obtained in strip 1, but the F1-scores of test strips 2–7 show an obvious drop compared to the results in Table 2.

Strip	1	2	3	4	5	6	7	8
P	0.862	0.668	0.722	0.817	0.798	0.737	0.731	0.861
R	1.000	1.000	1.000	0.967	0.925	0.985	0.999	1.000
F1	0.926	0.801	0.839	0.886	0.856	0.844	0.844	0.925

6. Conclusion

This study aims at achieving fine pixel-level PWD detection results, compared to the object-level detection used in most of the existing works. A novel PWD detection framework is proposed to deal with the difficulties when conducting pixel-level detection in complex forest landscapes, i.e., the need for a complete category system and the poor generalization. A preset category is difficult to set because there are numerous kinds of objects in complex forest landscape. Furthermore, the laborious labeling work is always necessary to obtain satisfying results. The poor generalization is common due to the spectral heterogeneity and the unseen objects, resulting in an obvious accuracy drop when detecting PWD in test images.

To avoid setting a category system, OCC is adopted as the basic principle of the proposed framework, which focuses on only the PWD samples. Because the traditional unbiased risk estimator favors the dominating negative objects (non-PWD) and the optimization of low-proportion PWD is ignored, the balanced unbiased detection risk estimator is proposed to make the optimization balanced. The hypothesis is that, if the PWD proportion value is small (Order of magnitude difference), it can be adjusted manually by copying PWD samples to make the risk of PWD and non-PWD objects balanced. To increase the generalization ability, a novel model consisting of three-dimensional (3D) convolutional layers and transformer blocks are constructed to extract more robust features. 3D convolution is more suitable for hyperspectral cube than 2D convolution due to the slide in spectral dimension, and the transformer blocks encourage the model to pay more attention to critical features and to correlate global features.

PWD detection experiments were conducted on eight UAV H² strips, with an imaging range of 400–1000 nm and a spatial resolution of 11 cm. Strip 1 was used to make training set and the remaining strips were used to make test set. The results showed that the proposed method can obtain fine pixel-level results with only a small number of labeled PWD samples (300 pixels in each strip, accounting for roughly 0.009 %). Furthermore, the proposed method has a better generalization ability than the existing pixel-level detection methods, which is a great benefit for practical usage. Considering the reduced labor cost and better generalization, this research has important practical significance for PWD control.

CRediT authorship contribution statement

Jingtao Li: Methodology, Software, Writing – original draft. **Xinyu Wang:** Conceptualization, Writing – review & editing, Project administration. **Hengwei Zhao:** Software, Investigation, Formal analysis. **Xin Hu:** Software, Validation, Visualization. **Yanfei Zhong:** Supervision, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The authors would like to give special thanks to the Shandong Provincial Institute of Land Surveying and Mapping for their assistance during the H² data collection. This work was supported by the National Natural Science Foundation of China under Grant Nos. 42071350 and 42101327, in part by the Fundamental Research Funds for the Central Universities under Grant No. 2042021kf0070, and in part by LIESMARS

Special Research Funding.

References

- Bellinger, C., Sharma, S., Japkowicz, N., 2012. One-class versus binary classification: Which and when?. In: 2012 11th International Conference on Machine Learning and Applications. IEEE, pp. 102–106.
- Bendale, A., Boult, T.E., 2016. Towards open set deep networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1563–1572.
- Cao, X., Zhou, F., Xu, L., Meng, D., Xu, Z., Paisley, J., 2018. Hyperspectral image classification with Markov random fields and a convolutional neural network. *IEEE Trans. Image Process.* 27 (5), 2354–2367.
- De Comité, F., Denis, F., Gilleron, R., Letouzey, F., 1999. Positive and unlabeled examples help learning. In: International Conference on Algorithmic Learning Theory. Springer, pp. 219–230.
- Deng, X., Tong, Z., Lan, Y., Huang, Z., 2020. Detection and location of dead trees with pine wilt disease based on deep learning and UAV remote sensing. *AgriEngineering* 2 (2), 294–307.
- Denis, F., Gilleron, R., Letouzey, F., 2005. Learning from positive and unlabeled examples. *Theor. Comput. Sci.* 348 (1), 70–83.
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., 2020. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv Prepr. arXiv2010.11929.
- Du Plessis, M.C., Niu, G., Sugiyama, M., 2015. Convex Formulation for Learning from Positive and Unlabeled Data, in: Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37, ICML'15. JMLR.org, pp. 1386–1394.
- Du Plessis, M.C., Niu, G., Sugiyama, M., 2014. Analysis of learning from positive and unlabeled data. *Adv. Neural Inf. Process. Syst.* 27, 703–711.
- Exelis, V.I.S., 2015. ENVI 5.3. Exelis VIS Boulder, CO, USA.
- Hao, Z., Huang, J., Li, X., Sun, H., Fang, G., 2022. A multi-point aggregation trend of the outbreak of pine wilt disease in China over the past 20 years. *For. Ecol. Manage.* 505, 119890 <https://doi.org/10.1016/j.foreco.2021.119890>.
- Haut, J.M., Paoletti, M.E., Plaza, J., Li, J., Plaza, A., 2018. Active learning with convolutional neural networks for hyperspectral image classification using a new bayesian approach. *IEEE Trans. Geosci. Remote Sens.* 56 (11), 6440–6461. <https://doi.org/10.1109/TGRS.2018.2838665>.
- Hirata, A., Nakamura, K., Nakao, K., Kominami, Y., Tanaka, N., Ohashi, H., Takano, K.T., Takeuchi, W., Matsui, T., Wong, W.O., 2017. Potential distribution of pine wilt disease under future climate change scenarios. *PLoS One* 12 (8), e0182837. <https://doi.org/10.1371/journal.pone.0182837>.
- Hu, X., Wang, X., Zhong, Y., Zhang, L., 2022. S3ANet: Spectral-spatial-scale attention network for end-to-end precise crop classification based on UAV-borne H₂ imagery. *ISPRS J. Photogramm. Remote Sens.* 183, 147–163. <https://doi.org/10.1016/j.isprsjprs.2021.10.014>.
- Hu, G., Zhu, Y., Wan, M., Bao, W., Zhang, Y., Liang, D., Yin, C., 2020. Detection of diseased pine trees in unmanned aerial vehicle images by using deep convolutional neural networks. *Geocarto Int.* 1–20.
- Ikegami, M., Jenkins, T.A.R., 2018. Estimate global risks of a forest disease under current and future climates using species distribution model and simple thermal model – Pine Wilt disease as a model case. *For. Ecol. Manage.* 409, 343–352. <https://doi.org/10.1016/j.foreco.2017.11.005>.
- Imambi, S., Prakash, K.B., Kanagachidambaresan, G.R., 2021. PyTorch. Programming with TensorFlow. Springer 87–104.
- Janocha, K., Czarnecki, W.M., 2017. On Loss Functions for Deep Neural Networks in Classification. On loss functions for deep neural networks in classification. arXiv Prepr. 1/2016 <https://doi.org/10.4467/20838476SI.16.004.6185>.
- Kiryu, R., Niu, G., du Plessis, M.C., Sugiyama, M., 2017. In: Positive-Unlabeled Learning With Non-Negative Risk Estimator, in. Curran Associates Inc., Red Hook, NY, USA, pp. 1674–1684.
- Lee, S.H., Cho, H.K., 2006. Detection of the pine trees damaged by pine wilt disease using high spatial remote sensing data. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* 36.
- Lei, L., Wang, X., Zhong, Y., Zhao, H., Hu, X., Luo, C., 2021. DOCC: Deep one-class crop classification via positive and unlabeled learning for multi-modal satellite imagery. *Int. J. Appl. Earth Obs. Geoinf.* 105, 102598. <https://doi.org/10.1016/j.jag.2021.102598>.
- Li, Y., Zhang, H., Shen, Q., 2017. Spectral-spatial classification of hyperspectral imagery with 3D convolutional neural network. *Remote Sens.* 9 (1), 67. <https://doi.org/10.3390/rs9010067>.
- Lim, E.T., Do, M.S., 2021. Pine Wilt Disease Detection Based on Deep Learning Using an Unmanned Aerial Vehicle. *KSCE J. Civ. Environ. Eng. Res.* 41, 317–325.
- Liu, Z., Zhong, Y., Wang, X., Shu, M., Zhang, L., 2021a. Unsupervised Deep Hyperspectral Video Target Tracking and High Spectral-Spatial-Temporal Resolution (H³) Benchmark Dataset. *IEEE Trans. Geosci. Remote Sens.*
- Liu, Zhenqi, Wang, X., Shu, M., Li, G., Sun, C., Liu, Ziying, Zhong, Y., 2021b. An anchor-free Siamese target tracking network for hyperspectral video. In: 2021 11th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS). IEEE, pp. 1–5.
- Matsuhashi, S., Hirata, A., Akiba, M., Nakamura, K., Oguro, M., Takano, K.T., Nakao, K., Hijioka, Y., Matsui, T., 2020. Developing a point process model for ecological risk assessment of pine wilt disease at multiple scales. *For. Ecol. Manage.* 463, 118010 <https://doi.org/10.1016/j.foreco.2020.118010>.
- Mäyrä, J., Keski-Saari, S., Kivinen, S., Tanhuapää, T., Hurskainen, P., Kullberg, P., Poikolainen, L., Viinikka, A., Tuominen, S., Kumpula, T., Vihervaara, P., 2021. Tree

- species classification from airborne hyperspectral and LiDAR data using 3D convolutional neural networks. *Remote Sens. Environ.* 256, 112322. <https://doi.org/10.1016/j.rse.2021.112322>.
- Mota, M., Vieira, P., 2008. Pine wilt disease: A worldwide threat to forest ecosystems, *Nematology*.
- Park, H.G., Yun, J.P., Kim, M.Y., Jeong, S.H., 2021. Multichannel Object Detection for Detecting Suspected Trees With Pine Wilt Disease Using Multispectral Drone Imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 14, 8350–8358.
- Perera, P., Oza, P., Patel, V.M., 2021. One-class classification: A survey. *arXiv Preprint arXiv2101.03064*.
- Piironen, R., Fassnacht, F.E., Heiskanen, J., Maeda, E., Mack, B., Pellikka, P., 2018. Invasive tree species detection in the Eastern Arc Mountains biodiversity hotspot using one class classification. *Remote Sens. Environ.* 218, 119–131. <https://doi.org/10.1016/j.rse.2018.09.018>.
- Rajasekharan, S.K., Lee, J.-H., Ravichandran, V., Lee, J., 2017. Assessments of iodoindoles and abamectin as inducers of methuosis in pinewood nematode, *Bursaphelochus xylophilus*. *Sci. Rep.* 7, 1–13.
- Ramaswamy, H., Scott, C., Tewari, A., 2016. Mixture proportion estimation via kernel embeddings of distributions. International Conference on Machine Learning. PMLR 2052–2060.
- Ruff, L., Vandermeulen, R., Goernitz, N., Deecke, L., Siddiqui, S.A., Binder, A., Müller, E., Kloft, M., 2018. Deep one-class classification. International Conference on Machine Learning. PMLR 4393–4402.
- Sabokrou, M., Fathy, M., Zhao, G., Adeli, E., 2020. Deep end-to-end one-class classifier. *IEEE Trans. neural networks Learn. Syst.* 32, 675–684.
- Schölkopf, B., Williamson, R.C., Smola, A.J., Shawe-Taylor, J., Platt, J.C., 1999. Support vector method for novelty detection., In: NIPS. Citeseer, pp. 582–588.
- Syifa, M., Park, S.-J., Lee, C.-W., 2020. Detection of the pine wilt disease tree candidates for drone remote sensing using artificial intelligence techniques. *Engineering* 6, 919–926.
- Wang, H., Lei, Z., Zhang, X., Zhou, B., Peng, J., 2016. Machine learning basics. *Deep Learn.* 98–164.
- Wu, B., Liang, A., Zhang, H., Zhu, T., Zou, Z., Yang, D., Tang, W., Li, J., Su, J., 2021. Application of conventional UAV-based high-throughput object detection to the early diagnosis of pine wilt disease by deep learning. *For. Ecol. Manage.* 486, 118986.
- Wu, W.B., Zhang, Z.B., Zheng, L.J., Han, C.Y., Wang, X.M., Xu, J., Wang, X.R., 2020. Research Progress on the Early Monitoring of Pine Wilt Disease Using Hyperspectral Techniques. *SENSORS* 20. <https://doi.org/10.3390/s20133729>.
- Yang, X., Ye, Y., Li, X., Lau, R.Y.K., Zhang, X., Huang, X., 2018. Hyperspectral image classification with deep learning models. *IEEE Trans. Geosci. Remote Sens.* 56, 5408–5423. <https://doi.org/10.1109/TGRS.2018.2815613>.
- Yu, R., Luo, Y., Li, H., Yang, L., Huang, H., Yu, L., Ren, L., 2021a. Three-Dimensional Convolutional Neural Network Model for Early Detection of Pine Wilt Disease Using UAV-Based Hyperspectral Images. *Remote Sens.* 13, 4065.
- Yu, R., Luo, Y., Zhou, Q., Zhang, X., Wu, D., Ren, L., 2021b. Early detection of pine wilt disease using deep learning algorithms and UAV-based multispectral imagery. *For. Ecol. Manage.* 497, 119493 <https://doi.org/10.1016/j.foreco.2021.119493>.
- Zhang, F., Du, B., Zhang, L., Zhang, L., 2016. Hierarchical feature learning with dropout k-means for hyperspectral image classification. *Neurocomputing* 187, 75–82.
- Zhao, H., Zhong, Y., Wang, X., Hu, X., Luo, C., Boitt, M., Piironen, R., Zhang, L., Heiskanen, J., Pellikka, P., 2022. Mapping the distribution of invasive tree species using deep one-class classification in the tropical montane landscape of Kenya. *ISPRS J. Photogramm. Remote Sens.* 187, 328–344.
- Zhong, Y., Hu, X., Luo, C., Wang, X., Zhao, J., Zhang, L., 2020. WHU-Hi: UAV-borne hyperspectral with high spatial resolution (H2) benchmark datasets and classifier for precise crop identification based on deep convolutional neural network with CRF. *Remote Sens. Environ.* 250, 112012.