symmetry

MDPI

Article

A Hybrid Vegetation Detection Framework: Integrating Vegetation Indices and Convolutional Neural Network

Wahidah Hashim [1], Lim Soon Eng [1], Gamal Alkawsi [2,*], Rozita Ismail [1], Ammar Ahmed Alkahtani [2], Sumayyah Dzulkifly [3], Yahia Baashar [2] and Azham Hussain [4]

[1] College of Computing and Informatics, Universiti Tenaga Nasional, Kajang 43000, Malaysia; Wahidah@uniten.edu.my (W.H.); limsooneng@live.com.my (L.S.E.); irozita@uniten.edu.my (R.I.)
[2] Institute of Sustainable Energy (ISE), Universiti Tenaga Nasional, Kajang 43000, Malaysia; ammar@uniten.edu.my (A.A.A.); yahia.baashar@uniten.edu.my (Y.B.)
[3] Computing Department, Faculty of Art, Computing and Creative Industry (FSKIK), Sultan Idris Education University, Tanjong Malim 35900, Malaysia; sumayyah.dzul@fskik.upsi.edu.my
[4] School of Computing, Universiti Utara Malaysia, Bukit Kayu Hitam 06010, Malaysia; azham.h@uum.edu.my
* Correspondence: gamal.abdulnaser@uniten.edu.my

Abstract: Vegetation inspection and monitoring is a time-consuming task. In the era of industrial revolution 4.0 (IR 4.0), unmanned aerial vehicles (UAV), commercially known as drones, are in demand, being adopted for vegetation inspection and monitoring activities. However, most off-the-shelf drones are least favoured by vegetation maintenance departments for on-site inspection due to limited spectral bands camera restricting advanced vegetation analysis. Most of these drones are normally equipped with a normal red, green, and blue (RGB) camera. Additional spectral bands are found to produce more accurate analysis during vegetation inspection, but at the cost of advanced camera functionalities, such as multispectral camera. Vegetation indices (VI) is a technique to maximize detection sensitivity related to vegetation characteristics while minimizing other factors which are not categorised otherwise. The emergence of machine learning has slowly influenced the existing vegetation analysis technique in order to improve detection accuracy. This study focuses on exploring VI techniques in identifying vegetation objects. The selected VIs investigated are Visible Atmospheric Resistant Index (VARI), Green Leaf Index (GLI), and Vegetation Index Green (VIgreen). The chosen machine learning technique is You Only Look Once (YOLO), which is a clever convolutional neural network (CNN) offering object detection in real time. The CNN model has a symmetrical structure along the direction of the tensor flow. Several series of data collection have been conducted at identified locations to obtain aerial images. The proposed hybrid methods were tested on captured aerial images to observe vegetation detection performance. Segmentation in image analysis is a process to divide the targeted pixels for further detection testing. Based on our findings, more than 70% of the vegetation objects in the images were accurately detected, which reduces the misdetection issue faced by previous VI techniques. On the other hand, hybrid segmentation methods perform best with the combination of VARI and YOLO at 84% detection accuracy.

Keywords: vegetation detection; vegetation indices; convolutional neural network; hybrid method

check for updates

Citation: Hashim, W.; Eng, L.S.; Alkawsi, G.; Ismail, R.; Alkahtani, A.A.; Dzulkifly, S.; Baashar, Y.; Hussain, A. A Hybrid Vegetation Detection Framework: Integrating Vegetation Indices and Convolutional Neural Network. *Symmetry* **2021**, *13*, 2190. https://doi.org/10.3390/sym13112190

Academic Editors: Chien-Hsing Chou and Yi-Zeng Hsieh

Received: 25 September 2021
Accepted: 26 October 2021
Published: 17 November 2021

1. Introduction

Vegetation inspection is a part of power energy companies' and environmental agencies' responsibility to maintain vegetation away from transmission and overhead power lines. The aim is to identify overgrown vegetation that has reached a specific clearance zone (see Figure 1) of the power structure. Once the vegetation reaches that zone, it will impact the transmission and distribution power lines [1].

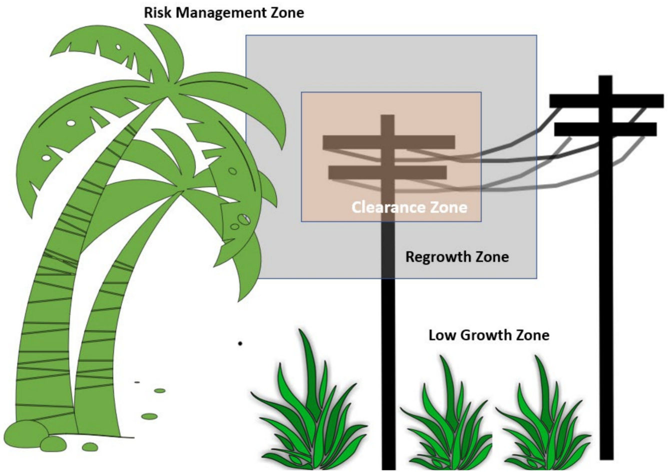


Figure 1. Example of the clearance zone for vegetation.

Researchers have proposed many remote sensing techniques to reduce the cost and shorten the consumption time of vegetation inspection [1]. Standard remote sensing techniques usually proposed for aerial image data collection are satellite or Unmanned Aerial Vehicle (UAV) equipped with a camera or specific sensors [2,3]. UAV provides advantages such as obtaining the latest information on target areas, low cost, and being able to fly at a lower altitude, as compared to other airborne aerial technologies [3–6]. Once land surface data are gathered, a computer system is needed to analyse the data to achieve particular objectives.

Vegetation indices (VI) are commonly introduced as data analysis techniques to differentiate between vegetation and non-vegetation objects on the land surface. VIs is a simple and effective algorithm for quantitative and qualitative evaluation of vegetation coverage, vegetative vigour test, and plant growth dynamics using the remote sensing method [7]. Different VI techniques serves different purposes and different data requirements for their analyses. Among other examples of VI techniques are Visible Atmospheric Resistant Index (VARI), Green Leaf Index (GLI), and Vegetation Index Green (VIgreen). These techniques require only Red, Blue, and Green (RGB) channel data for analysis [8,9]. Most related research has proven that these VI techniques show accurate results, especially for plantation field data. However, in urban areas that consist of many object categories, error judgment is present for non-vegetation objects that are green in colour, leading to inaccurate object identification. An experiment has proven that the result of VIgreen and VARI could misclassify some non-vegetation objects as vegetation because the values given are in the range of the vegetation category [10].

UAV is gradually adopted for performing the inspection in Malaysia. However, according to [11], it was found that UAV technologies were very demanding in power transmission and distribution industries due to their inspection efficiency improvement. The limitation of the vegetation recognition technique for remote sensing using RGB images requires more attention and research. The aerial image analysis technique is critical for a computer to process the data gathered from an UAV. Many researchers have introduced VI to detect vegetation health and changes [12]. However, most of the research mainly focuses on the forest and plantation field, where object classifications are minimal.

Another challenge in this VI aerial image analysis technique is that it requires a multispectral or modified camera to capture the near-infrared (NIR) channel. NIR channels are not visible by normal human vision and use a typical consumer camera [13]. Mark C. Dustin (2015) experimented with a low-cost UAV to capture aerial images and post-process these images with the VI technique, which required both RGB and NIR colour band characteristics for parks' monitoring data [14]. The main reason for combining RGB and NIR band techniques is to improve the sensitivity of detecting green vegetation [7]. The findings from the experiment show inaccurate detection of non-vegetation objects due to camera limitation in the ability of white balance setting. The white balance setting is essential to ensure colours in the images are as accurate as possible.

In [15], the researchers experimented using multispectral and RGB remote sensing images to detect treetops on the ground surface. The proposed processing method was combined with a fully convolution neural network (FCN) and two unsupervised treetop detectors. The result found that RGB achieved less than 10% lower accuracy, as compared to multispectral images. Nonetheless, both experiments proved the limitation in urban areas by detecting non-vegetation objects as tree objects.

In recent years, few researchers have focused on using only consumer-based equipment and visible band data for vegetation detection [16]. It is understood that RGB-based camera UAVs are cheaper, when compared to multispectral camera UAVs [17]. The choice of selecting an RGB-based camera UAV may be preferable due to the lower cost acquired. This is why RGB-based camera UAVs with proper aerial image analysis techniques are considered in our research to achieve a cost-effective and accurate vegetation inspection. This study aims at exploring VI techniques in identifying vegetation objects. Hybrid methods between VI and machine learning will be developed to overcome the limitation of existing VI techniques. The proposed hybrid approaches will be tested on captured aerial images to observe vegetation detection performance. As a result of the limitations discovered in the initial proposed hybrid methods, a hybrid segmentation method is then introduced.

Section 2 presents the background of this study, while Section 3 discusses the related studies. Section 4 introduces the method, followed by Section 5, which highlights and discusses the study results. Section 6 presents the implications of the findings; Section 7 presents the limitations and potential future research directions. This article is concluded in Section 8.

2. Background

This section provides an overview of the YOLO model and vegetation indices.

2.1. You Only Look Once (YOLO)

The YOLO model does not require high computation power to execute the training process and can provide fast detection. It has a straightforward layer of detection. Additionally, the YOLO model is easy to construct and able to train on full images directly. The architecture for original YOLO consists of 24 convolutional layers, followed by two fully connected layers. By continuously learning image data, convolutional neural networks may extract features that can discriminate visual properties, similar to the delicate symmetry of a human brain learning. As for the YOLO detection system, it has three processing steps. First, it resizes the input image into a suitable size for YOLO detection. Then, it runs a single convolutional network with the resized image. Lastly, the detection object confidence result that meets the thresholds shows up in the final image.

In a technical report written by [18], the new version known as YOLO version 3 (YOLOv3) was introduced. The focused improvement on this version was to achieve higher accuracy with several new hidden layers in the model. Compared to the previous version, this version added the feature pyramid network in the model. It is used to predict class, bounding box, and objects better in the image. The feature classification approach named Darknet-53 was also implemented to improve the feature extraction process. Darknet-53 starts to perform convolution from 256 × 256 frame size until the last of 8 × 8. In between, residual layers are combined, and some are repeated. According to the study, an experiment was conducted to compare YOLOv3 with other common detection models. The result found that YOLOv3 consisted of an additional layer above, which was able to improve the accuracy of the bounding box on the detected objects and still had faster detection, compared to the common detection model.

2.2. Vegetation Indices

Vegetation indices is a typical image analysis technique for remote sensing of vegetation. VIs are divided into two categories: multispectral VI and visible band VI. Multispectral VI requires a near-infrared (NIR) channel with a combination of RGB channels to calculate the index. In contrast to visible band VI, it only requires RGB channel value to do so. In this section, only selected VI will be discussed as per the mechanism relevant to our study.

2.2.1. Green Leaf Index (GLI)

GLI was recommended by [9] in identifying the grazing impacts of aerial photography on wheat. Their research aimed to estimate wheat cover on the ground level by using the aerial view of some RGB images. During the experiment, they manipulated the use of an RGB camera and position in nadir angle to capture ground surface as a prototype aerial photography. GLI was developed for this experiment to differentiate between the green leaves from the non-living object. The values of GLI were returned between the range of −1 and +1. Each positive and negative value represents vegetation and non-living objects, respectively. If the pixels produce GLI value in a positive number, it represents green leaves; otherwise, it is assumed as soil or non-living objects. Another research found that GLI is best used to determine rice leaf chlorophyll due to the equation sensitivity to a greenish leaf [19]. However, in the vegetation field analysis experiment, it proved that the effectiveness of GLI was on identifying exposed soil [20]:

$$\mathrm{GLI} = (2\mathrm{Green} - \mathrm{Red} - \mathrm{Blue}) / (2\mathrm{Green} + \mathrm{Red} + \mathrm{Blue}) \tag{1}$$

2.2.2. Vegetation Indices Green (VIgreen)

VIgreen was introduced by [8], and it only uses the red and green bands to interpret various ground cover, and allows the identification of green vegetation and non-vegetation based on the value of the equation. Tucker C. J. (1979) used a similar equation on performing vegetation monitoring based on satellite images with specific values, in comparison between photographic infrared (IR) with a red and green combination [21]. Although VIgreen was adopted in the work, the research was more focused on IR with red combination, and it proved that the IR performed better than the green and red combination on vegetation monitoring detection using satellite images. In contrast, Ref. [9] used Equation (2) to develop another VI, which was known as VARI. VIgreen was also known as Green-Red Vegetation Index (GRVI), Normalized Green-Red Difference Index (NGRDI), and Nitrogen Reflectance Index (NRI) [22–24].

$$\mathrm{VIgreen} = (\mathrm{Green} - \mathrm{Red}) / (\mathrm{Green} + \mathrm{Red}) \tag{2}$$

2.2.3. Visible Atmospheric Resistant Index (VARI)

VARI is mainly used to estimate vegetation fraction with minimal sensitivity to atmospheric effect [8,25,26]. This VI is an enhancement of VIgreen by adding the blue channel in Equation (3). According to [27], the additional blue channel in the equation reduced the atmospheric effects by subtracting the blue band. The atmospheric effects became less sensitive, and the result of estimation vegetation fraction was able to achieve less than ten per cent of error [8,26]. According to VIs comparison studies by [18], VARI was considered a reliable measure of vegetation field health for RGB-based images. It represented the major features of the multispectral VI Normalized Difference Vegetation Index.

$$\mathrm{VARI} = (\mathrm{Green} - \mathrm{Red}) / (\mathrm{Green} + \mathrm{Red} - \mathrm{Blue}) \tag{3}$$

3. Related Studies

The UAV has become one of the most common tools used by many industries for a specific task. The UAV can provide low operational costs, while capturing high-quality images [28,29]. With a proper image processing technique, UAV data processing could provide more advantages on remote sensing tasks. Image processing requires a specific algorithm to process input data for further analysis according to the task objective. For example, vegetation indices (VI) are mainly used for vegetation change, vegetation health, and vegetation detection [30,31]. Other than VI, machine learning can also focus on the innovation of machine vision on vegetation monitoring.

The literature shows that research focuses on object detection using machine learning to incorporate machine vision in identifying vegetation objects. For instance, a work lead by [32] proposed a deep convolutional neural network (CNN) for palm tree detection from remote sensing data. The CNN model used in the study is AlexNet. Based on their findings, it was proven that the proposed method could detect palm trees with acceptable performance. Research by [33] suggested a spectral-spatial approach based on a neural network to detect and count tree crowns from images captured by an UAV. The proposed model used an extreme learning algorithm to identify the tree crown from their image understudies. Other than using UAV, the authors of [11] performed an experiment using fully convolution networks (FCNN) to detect roadside vegetation from RGB imagery. This experiment was conducted using RGB images captured from a camcorder installed on a moving vehicle. The result of the experiment proved that the proposed network performed better in interpreting vegetation and non-vegetation objects than the two selected VIs, as well as another well-known machine learning called the support vector machine (SVM).

An experiment using machine learning for vegetation monitoring along a power line corridor was experimented by [1]. One of the primary motivations for executing such work was the advantage of machine vision technology to monitor vegetation at low cost for efficient power line monitoring. The research used the pulse coupled neural network to perform tree detection and classification from multispectral imagery and LiDAR data. The test was included in another machine learning technique known as SVM to perform more advanced tree species classifications. Another similar method was introduced by [34]) using the combination of imagery and LiDAR with self-supervised deep learning neural networks algorithm. Nonetheless, the proposed method had a challenge in detecting smaller trees. However, the authors commented that the existing technique required further improvement to detect different tree conditions.

Machine learning is not only limited to vegetation-related data. According to [35], the unsupervised machine learning algorithm for machine vision of plantations yielded a better-quality estimation of the production value of the vegetation. A review by [36] noted several deep learning models on agriculture that can perform better than the conventional image processing analysis in the agricultural domain. Deep learning is an extended version of the machine learning model that allows the machine to learn through multiple levels of analysis. Our study has to consider using one of the deep learning models founded by [18].

Few studies applied remote sensing vegetation using UAV. For example, Ref. [37] suggested a method for detecting and counting banana plants using RGB images captured from an UAV. These methods included Linear Contrast Stretch, Synthetic Color Transform, and Triangular Greenness Index (TGI). The method used a convolution neural network with three combinations of image processing. Their research was to merge the result from three different image processing method combinations and test with a different altitude of the dataset. The result showed that 87% plants were correctly detected and above in all altitudes. Neupane et al. (2019) explored a method with one VI and two image processing, which is in contrast to our research. In our work, we will compare three different VIs and one focused machine learning [37]. A work by [24] conducted an experiment using multispectral and RGB UAV aerial image for detecting disease symptoms in sugarcane fields. This experiment used the multispectral VI algorithm that was part of NDVI and RGB VI in GLI, VIgreen, and TGI. The results found that multispectral VI had the best detection of symptoms; however, the RGB VI, especially VIgreen, had the ability to detect the symptoms with an acceptable result of showing symptoms based on the value difference and also offered a less-costly solution. Recently, research conducted by [38] performed a comparison of the combination of VI and machine learning to interpret rice lodging also based on RGB images obtained by an UAV. The focused VI were ExG, ExR, ExG−ExR and the machine learning ones were AlexNet and SegNet. The goal of the research was to use these combination methods to identify rice paddy, rice lodging, and other objects such as roads and bare land. The result showed that the combination of ExG−ExR with AlexNet had the best accuracy in their case study.

4. Method

This section provides details about the proposed vegetation detection system. The detection system was divided into two parts. The first part of detection involves converting original RGB images into VI results and removing pixels with VI value below 0 for HSM. The second part performs tree detection from the processed images using YOLO.

4.1. Data Collection

The data were collected from two sites—the first was located at University Tenaga Nasional, Putrajaya Campus (UNITEN), and the second at Taman Semenyih Indah, Selangor (TSI). Both were done during good sunny weather with acceptable wind speed conditions. The first flight was executed at an empty field in UNITEN, as shown in Figure 2. The second flight was launched at an empty field at TSI, as shown in Figure 3.

Spark from the DJI drone brand was selected as COTS UAV. The reason for choosing DJI Spark was that it is equipped with a 12-megapixels RGB sensor camera. The video quality that could be captured is up to 30 frames per second (fps) in high-definition 1920 width and 1080 height frame size. The DJI Spark camera can rotate into a nadir view from the air, due to the two-axis mechanical gimbal connected between the camera and the UAV. However, autopilot for image mapping is not available for that model, which is considered a limitation of that drone.

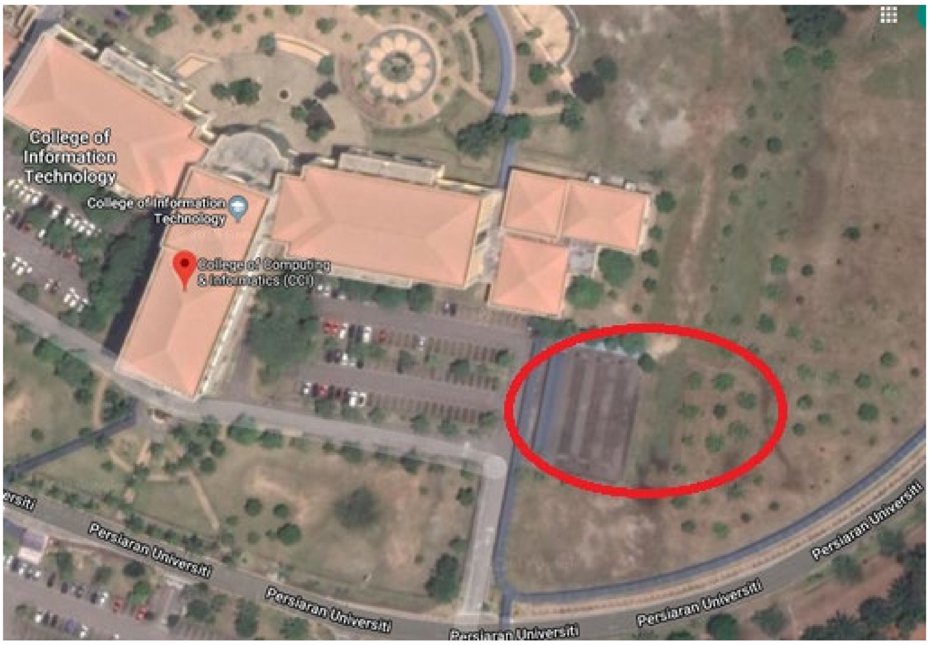


Figure 2. Selected study site 1—UNITEN.

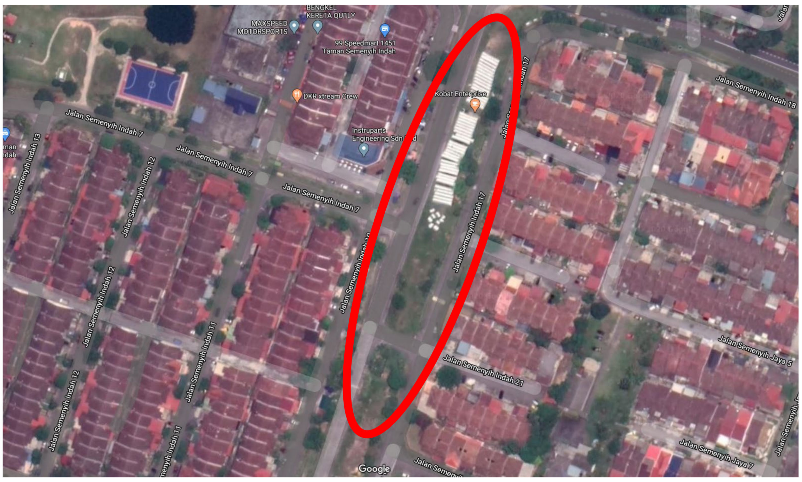


Figure 3. Selected study site 2—TSI.

After the video was gathered from the UAV, a few pre-processing steps were required to convert raw videos footage into VI images. First, each frame was extracted from the video and saved into a Joint Photographic Experts Group (JPEG) image format file that consisted of a RGB band. A total of 13,168 images were selected and extracted from the video footage on both study sites. There was a total of 5600 images from footage captured at UNITEN and a total of 7568 images in the footage captured at TSI. All images were from the nadir view of the study sites. To ensure both sites were in an equal amount in the dataset, an additional step of random picking was required to be performed on the data from TSI. These random pick processes were selected using a build-in feature in Matlab. Each of these images was assigned a sequence of numbers randomly from 1–5600.

4.2. Dataset

Overall, a total of 11,200 images were chosen with an equal number of the dataset for both study areas. The collected images from both study sites were classified as below:

UNITEN site: Trees are not crowded together, and most of the trees in this study site are far from each other.

TSI site: Trees are planted randomly, not appropriately arranged, with additional challenges on site (such as construction material) and multiple types of plantation around.

4.3. Data Processing

The processing stage was divided into two parts. The first involved inserting the images into an open-source application named YOLO marker for image-labelling. An example of the interface is shown in Figure 4. This step labelled focused objects such as trees and vegetation according to the research requirement. In our experiment, not all of the images contained the focused objects; part of the images consisted of a ground surface with different objects such as greenfield, tarmac, rooftop, and other non-vegetation objects. Simulating the non-vegetation situation aimed to help the YOLO practice to differentiate the non-vegetation from vegetation images. Once each image was labelled, the tools then proceeded to generate a text file for the YOLO training to identify the focused objects. Another parallel action was to use the images generated from pre-processing steps of converting them into VI images. An open-source geographical information system such as QGIS was used. A built-in function known as raster calculation was inserted to calculate the VI result. A python script was coded to automate the process of RGB images into VI images.

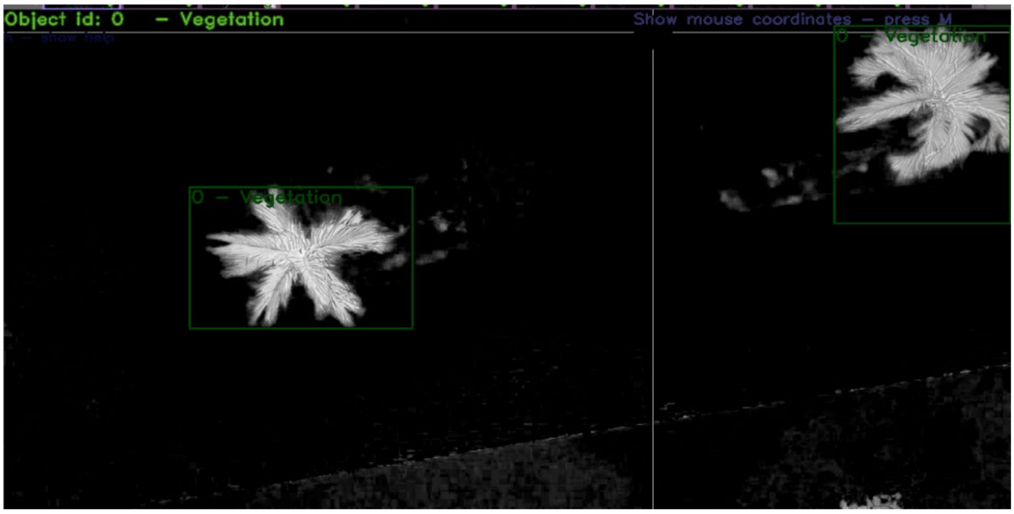


Figure 4. Example of the YOLO marker interface with labelled VI images.

The second process is to combine the VI Image with the labelling text file. The purpose of performing this combination was to reduce the amount of image labelling. Since the focused objects for each image were maintained the same and the labelling was required to be performed manually, it was better to use RGB image for labelling, then duplicating the generated text file to each of the VI images. This was due to a better view produced by the RGB images. This step could reduce the time taken to label each of the VI images manually.

Additionally, it provides the equality of labelling for each of the proposed methods. Unfortunately, the YOLO marker was unable to read the VI images generated from QGIS. To resolve this issue, an extra step is required, where the 1-band grayscale images generated were converted from QGIS and then converted back to 3-band JPEG grayscale images using Adobe Lightroom. These steps did not impact the quality of the images but still maintained the original grayscale pixel.

4.4. Image Processing

The image processing processes were developed and tested at different stages to improve accuracy and reduce the incorrect classification object. Figure 5 shows that the areas highlighted in yellow are the enhancements from the initial model with additional segmentation, applied mainly to eliminate most of the non-vegetation object pixels from the initial images. The other reason is to avoid any confusion about machine prediction before using machine learning detection. This additional process requires modification in Darknet, such as image segmentation based on VI results. According to [9], the negative value from GLI indicates non-living objects.

In this case, it is necessary to acquire more in-depth learning of Darknet coding to implement VI calculation function and segmentation function within the process. Darknet was developed using C programming language with an additional plugin, such as OpenCV. As a result, it gives the advantage of implementing the RGB to VI conversion directly into Darknet instead of using GIS to calculate, as in the above experiment. After converting RGB to VI, each pixel value was then stored to perform pixel segmentation by eliminating non-vegetation objects pixels based on VI value below 0. In contrast, VI value above 0 remains within the RGB value and is displayed as usual. Figure 6 is one of many example images captured from UAV with the output result of selected VI pixel segmentation. Next, each segmented image was rendered into an image file to input into the YOLO machine learning training and detection algorithm.

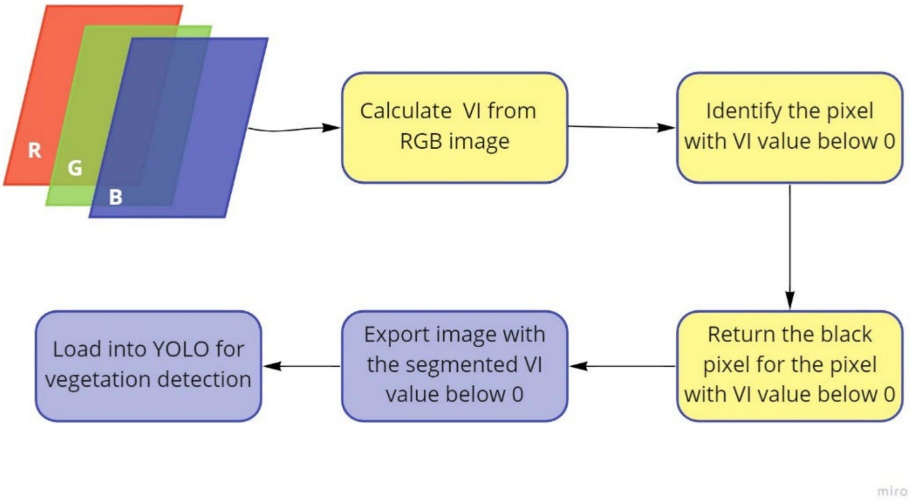

R
G
B
Calculate VI from RGB image
Identify the pixel with VI value below 0
Return the black pixel for the pixel with VI value below 0
Export image with the segmented VI value below 0
Load into YOLO for vegetation detection
miro

Figure 5. A flowchart for processing images using the proposed method.

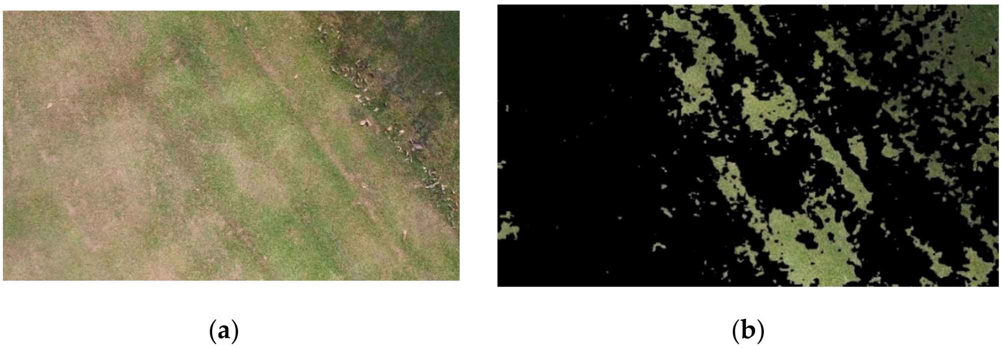


Figure 6. An original RGB image captured using UAV (**a**) and VI segmented images (**b**).

4.5. Training

Before the training began, the data were separated into two groups, classified as training and testing data. The data were then split into 70% for the training group, while the remaining 30% fell into the testing group. Overall, 7840 data out of 11,200 were grouped into the training category. Each study site selects a total of 3920 images as a training group. The remaining 3360 data were randomly chosen for the testing group, and both study sites were split equally, with a total of 1680 images in each testing group. All random pick process was executed using the same process, whereby 5600 images from TSI dataset were produced. All data splitting processes for each of the VI images followed the same steps.

After all the data were pre-processed, the training group data were then ready for YOLO training. The YOLO version 3 (YOLOv3) introduced by [18] was selected for this experiment. The focused improvement on this version was to achieve higher accuracy with several new hidden layers in the model. Compared to the previous version, this version added the feature pyramid network in the model, used to predict class, bounding box, and objects better in the image. The feature classification approach named Darknet-53 was also implemented to improve the feature extraction process. Darknet-53 starts to perform convolution from 256 × 256 frame size until the last of 8 × 8 (See Figure 7). In between, residual layers are combined and some are repeated.

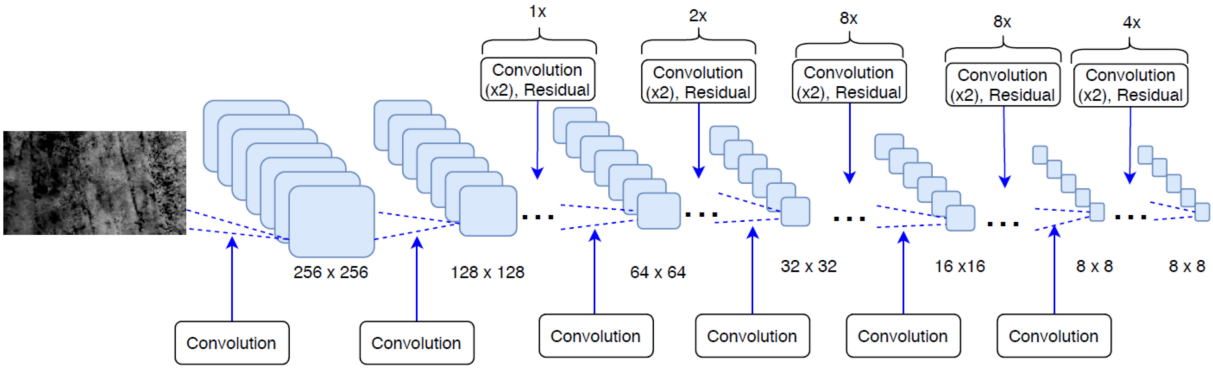

1x
2x
8x
8x
4x
Convolution (x2), Residual
Convolution (x2), Residual
Convolution (x2), Residual
Convolution (x2), Residual
Convolution (x2), Residual
256 x 256
128 x 128
64 x 64
32 x 32
16 x16
8 x 8
8 x 8
Convolution
Convolution
Convolution
Convolution
Convolution
Convolution

Figure 7. The simplified version of the architecture of YOLOv3 (Darknet-53) [18].

Darknet was required to conduct training and testing. Each VI dataset was trained through 3000 iterations within 64 batches. Most of the HM training convergence was around 300 iterations. The computer machine specification for executing the training was equipped with Intel (California, CA, USA) i7 8700 six-core processor unit (CPU), Nvidia (California, CA, USA) RTX 2070 graphics processing unit (GPU), and 16GB GDDR4 random access memory (RAM); for more information, see Table 1. Each training took 1 to 1$^1/_2$ h for 1000 iterations using a computer, with specifications as declared above. Therefore, each method required at least 5 h to reach the target iteration. Detection through video consumed the most processing power, compared to performing detection with a picture. Therefore, this research converted the video into a frame image before performing detection through the proposed method.

Table 1. Main training specifications and configurations.

Description	Specification
Backbone model	YOLO version 3
Number of Iteration	5000 Iterations
Learning Rate	0.001%
Input Image Size	1920 × 1080
Total sample	11,200 Images
Number of training samples	7840 Images
Number of testing samples	3360 Images
Batch	64
Subdivision	32
GPU specification	NVIDIA GeForce RTX 2070 8GB GDDR6 VRAM (x1)
CPU specification	Intel Core i7 8700 3.20 GHz, 6-Core Processor
RAM capacity	16 GB

4.6. Analysis

Two analysis methods were utilised at this stage. First, a convergence graph was used to compare each of the proposed methods' training progress. A technique was needed to analyse the machine learning performance of each method during the training process. A convergence graph was based on average loss value during the machine learning training stage:

$$\text{Loss} = \text{PB}_{(\text{x+y})} + \text{PB}_{(\text{w+h})} + \text{PB}_{(\text{CS})} + \text{PB}^{\text{no}}_{(\text{CS})} + \text{PC}_{(\text{Class})}$$

The equation above is the simplified version calculation of average loss introduced by [39]. During the machine learning training stage, an average loss value performs continuous calculation in every complete iteration. It determines the training result performance. There are five main terms in this average loss equation: $\text{PB}_{(\text{x+y})}$ and $\text{PB}_{(\text{w+h})}$ denote the loss of centroid coordinate in size and bounding box between the predicted and expected, respectively. Next, $\text{PB}_{(\text{CS})}$ and $\text{PB}^{\text{no}}_{(\text{CS})}$ represent the loss between the object's pixel and the non-object's pixel. The last term $\text{PC}_{(\text{Class})}$ represent the loss value on object classification. Summing up all the terms will determine the total average loss of current training.

Next, the result of the experiment was analysed using a confusion matrix. The study [40] used the confusion matrix method by comparing results between the manual reference and the CNN prediction to identify the accuracy of the machine's vision performance. In our experiment, the confusion matrix was applied to determine the accuracy of each hybrid method. The confusion matrix can classify prediction results into three categories, namely true positive (TP), true-negative (TN), and false-negative (FN). To determine each of the evaluations, it is compulsory to match the two criteria (see Tables 2 and 3). The first criterion is vegetation detection, which refers to two possible conditions on the image. Since not all testing images consist of vegetation and trees, the possible condition is either available or not available. The second criterion is the number of bounding boxes that detect all the expected tree objects. This criterion was able to observe the total bounding box that matches the expected tree objects and all bounding boxes assigned to the tree objects. To determine each of the confusion matrix evaluations, two criteria must be fulfilled to determine the TP. In contrast, one criterion is required for FN. On the other hand, TN is determined when trees are not expected or predicted.

Table 2. Requirement of confusion matrix evaluation.

Confusion Matrix Evaluation	Requirement
TP	When trees are correctly expected and predicted.
FN	When trees are expected and predicted, and only one criteria is fulfilled. When trees are not expected, but are predicted. When trees are expected, but are not predicted. When trees are expected and predicted, but the boundary box is unable to locate them correctly. When trees are expected and predicted, but not all expected trees are detected.
TN	When trees are not expected or predicted.

Table 3. Confusion matrix table.

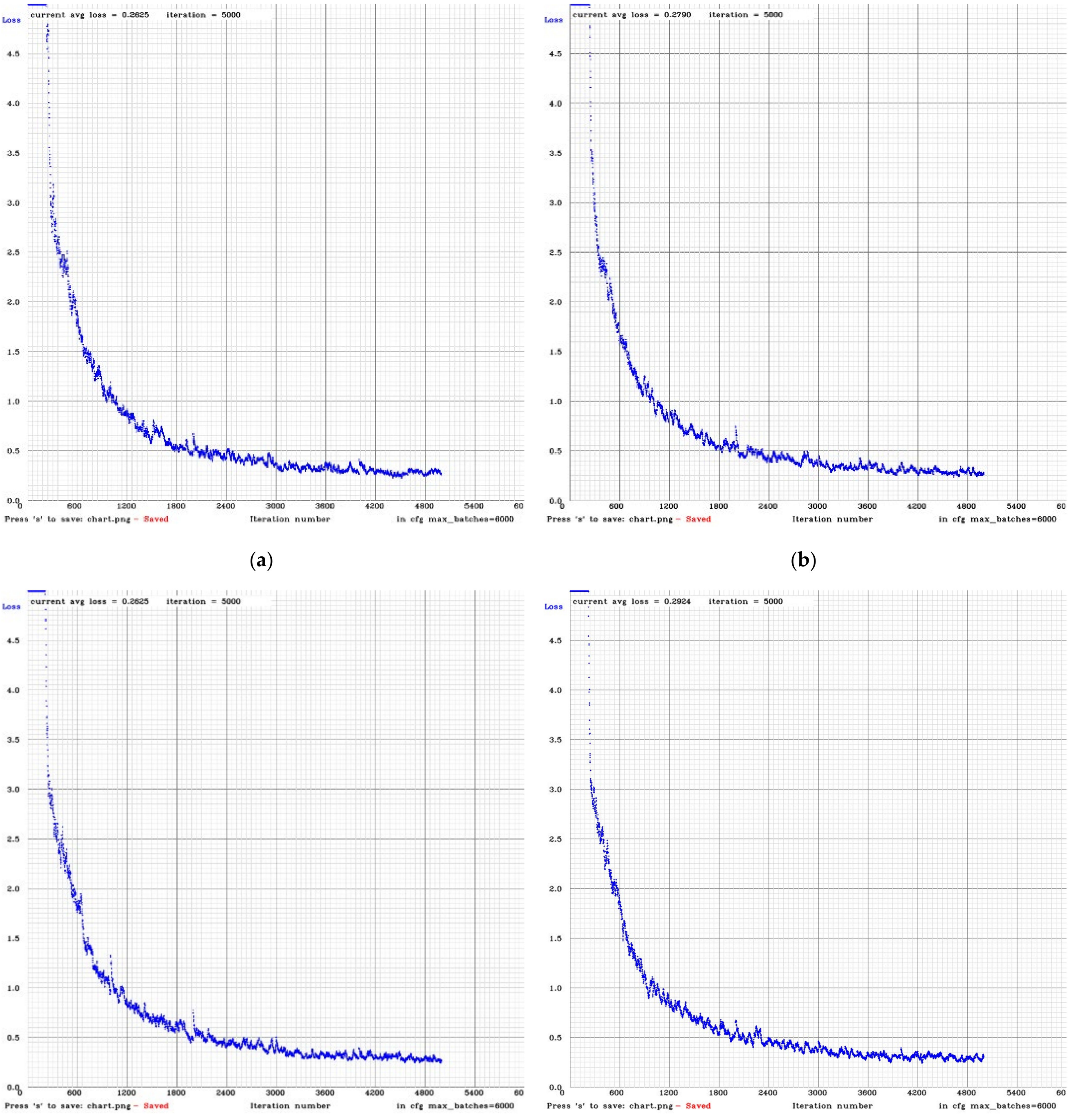
Total Image = 3360		**Predicted**	
		No	**Yes**
Actual	No	TN	
	Yes	FN	TP

5. Results

This section compares the experiment result using a different hybrid method and discusses the strengths and limitations. For better comparison, a basic method result is added—it directly using RGB images and performs detection through a YOLO machine learning.

5.1. Convergence Graph Comparison

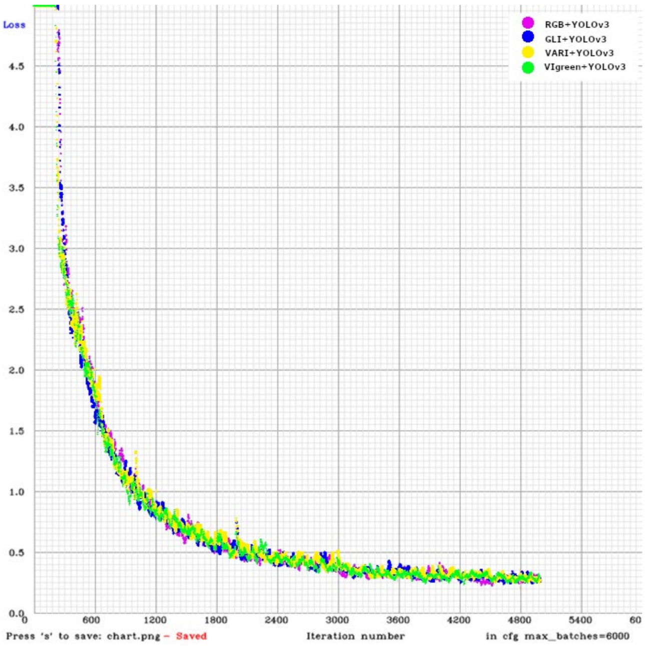
As mentioned earlier, each of the trainings conducted in Darknet produce a value calculated with a specific formula with a graph, as shown in Figure 8, to observe the performance in every iteration. The VARI + YOLOv3 method achieved an equal average loss through the normal RGB + YOLOv3 method with the lowest average loss value in the experiment, followed by the GLI + YOLOv3 method with a slightly different result in terms of value. The method with the highest average loss is VIgreen + YOLOv3. Even though the average loss decreased, the differences between every 1000 iteration (see Table 4) show a slightly decreased value towards the end of the training, which mean the training result with the lowest average loss has potential for better prediction and detection because the model learned more and was able to gather more accurate information. Therefore, it is worth studying the extent of duration training for this experiment.



(**a**) (**b**)

(**c**) (**d**)

Figure 8. *Cont.*



(**e**)

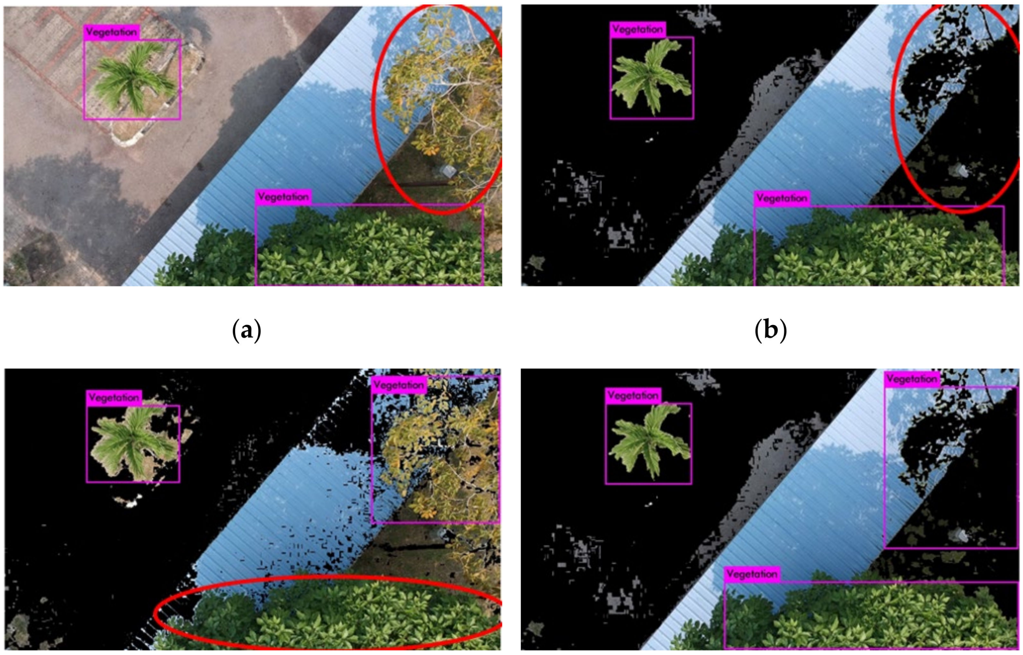
Figure 8. (**a**) Average loss graph for the RGB + YOLOv3 method; (**b**) Average loss graph for the GLI + YOLOv3 (Segmentation) method; (**c**) Average loss graph for the VARI + YOLOv3 (Segmentation) method; (**d**) Average loss graph for the VIgreen + YOLOv3 (Segmentation) method; (**e**) Average loss graph comparison.

Table 4. The average loss of the proposed methods.

Methods	RGB + YOLOv3	GLI + YOLOv3	VARI + YOLOv3	VIgreen + YOLOv3
1000 Iteration	1.0795	1.1015	0.971	1.0485
Different Between 1000 and 2000 Iteration	−0.5869	−0.5601	−0.5069	−0.4997
2000 Iteration	0.4926	0.541	0.4642	0.548
Different Between 2000 and 3000 Iteration	−0.1292	−0.1932	−0.0339	0.1707
3000 Iteration	0.3634	0.3482	0.4303	0.3781
Different Between 3000 and 4000 Iteration	−0.0983	−0.0468	−0.1445	−0.0613
4000 Iteration	0.2651	0.3014	0.2858	0.3168
Different Between 4000 and 5000 Iteration	−0.0026	−0.0224	−0.0233	−0.0244
5000 Iteration	0.2625	0.279	0.2625	0.2924

5.2. Confusion Matrix Analysis

One of the experiment images used to test the YOLO detection process is presented in Figure 8; the square pink boundary box is the prediction of the YOLO method. In contrast, the red circle is the tree that is not predicted by YOLO. The black area represents a below-negative VI value that was removed during the segmentation stage. As clearly shown in Figure 8, VIgreen segmentation with YOLO is able to identify all of the expected trees from the image. In this condition, this is categorized as TP, as it accurately located all expected trees and labelled the boundary box. The original RGB only detected two trees in each image. The GLI segmentation method was unable to detect the tree at the bottom right of the image. RGB with YOLO and VARI segmentation was not able to detect a yellowish tree at the top right side. As shown in Figure 9, there is another difference in how the VIs segmentation determines a yellowish tree. GLI still maintains the yellowish tree on the top right corner, but VIgreen and VARI were segmented and fully removed that particular tree due to its negative VI value. Since the segmentation stage removed the pixel with a negative VI value, the yellowish tree was removed from the result of VARI and GLI.

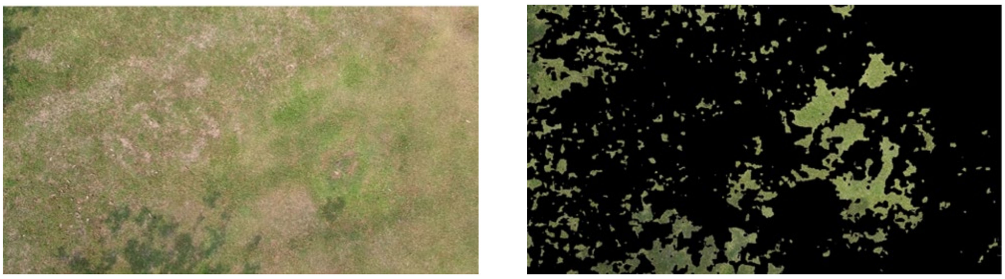


(**a**) (**b**)

(**c**) (**d**)

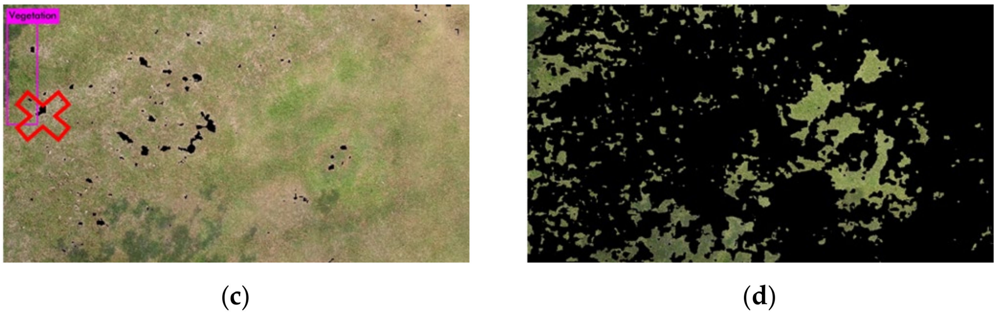
Figure 9. Image from UNITEN showing objects such as tarmac road, roof, and trees. (**a**); RGB = FN (1 missing tree); (**b**) VARI Seg = FN (1 missing tree); (**c**) GLI Seg = FN (1 missing tree); (**d**) VIGREEN Seg = TP (All trees correctly detected).

The second example is an empty green field without any trees surrounded it. No tree is expected to be detected; hence, there should not be any boundary box in the image. The comparison between the four figures in Figure 10 clearly shows that none of the boundary boxes exist in the original RGB and in two of the other methods known as VARI and VIgreen. Nonetheless, the GLI segmentation method displays one boundary box in the top left corner. It was an incorrect prediction. The part of the boundary covered area had a small tree's shadow at the edge, on the left. However, this condition was counted as FN because it was not expected to detect a tree or the boundary box. Other than that, the above comparison also clearly shows VARI and VIgreen able to segment the exposed sand area. This result proved that VARI and VIgreen are sensitive to non-vegetation areas and are able to define with correct negative value as the result. In contrast, GLI remains the sand-exposed area with a positive value.



(**a**) (**b**)

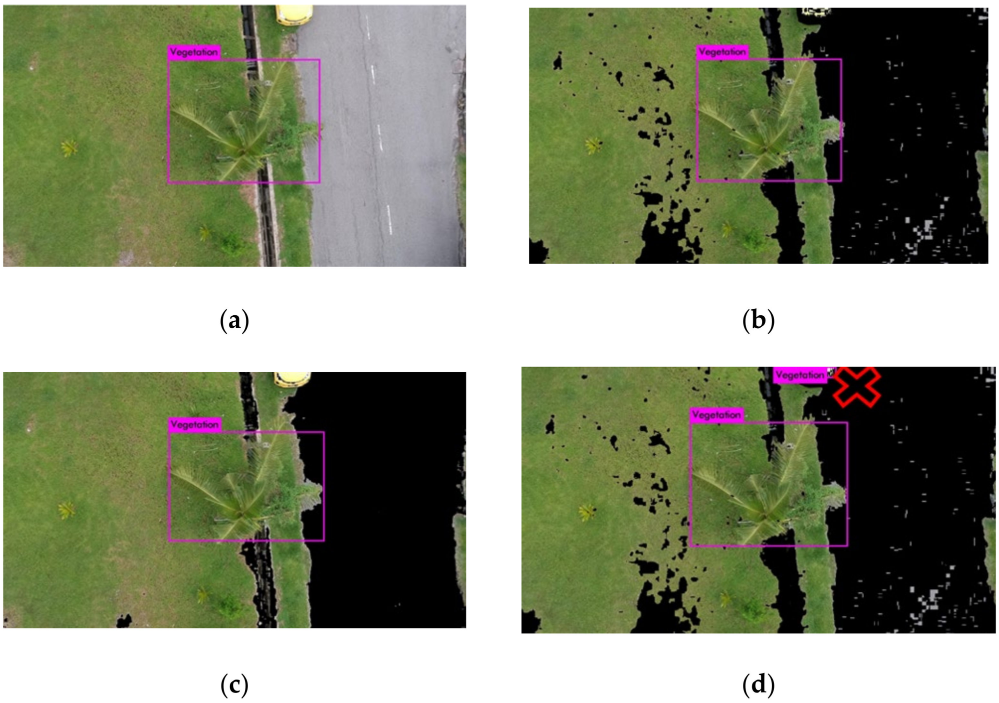
Figure 10. *Cont.*



(**c**) (**d**)

Figure 10. Image from UNITEN with view of a green field. (**a**); RGB = TN (correct with no tree detected); (**b**) VARI Seg = TN (correct with no tree detected); (**c**) GLI Seg = FN (Wrong detection); (**d**) VIGREEN Seg = TN (correct with no tree detected).

The third example (see Figure 11) is an image captured next to a residential area with a tarmac road on the right and a tree located in the middle of the frame and a car parked at the top of the image. The results show that all the methods detected the expected tree in the middle of the image. However, the VIgreen segmentation method detected another “vegetation” on top of the parked car. This is counted as an incorrect prediction. Therefore, this incorrect detection is categorized as FN because it did not achieve the expected amount of bounding box criterion as a result of more than one boundary box. The rest of the results are categorized as TP, as they meet both of the criteria correctly, except for the VIgreen segmentation method. Based on this example, all three methods are able to remove most of the non-vegetation objects, especially the tarmac roads.



(**a**) (**b**)

(**c**) (**d**)

Figure 11. Image from TSI with the view of tarmac and a small green field on the side of the road. (**a**); RGB = TP (all trees correctly detected); (**b**) VARI Seg = TP (all trees correctly detected); (**c**) GLI Seg = TP (all trees correctly detected); (**d**) VIGREEN Seg = FN (detected non-veg object).

The fourth example was obtained by TSI, but this example (See Figure 12) in respect to the algorithm detecting two small size trees. There were two small trees located on the right side of the image with a yellowish tree and one with green leaves. Based on the result, the original RGB method and VARI segmentation method managed to identify both small trees correctly. However, the GLI and VIgreen segmentation methods were unable to detect the yellowish trees. In addition, VIgreen and VARI segmentation had eliminated most of the yellowish trees from the image. Overall, both the original RGB method and VARI segmentation method are categorized as TP because they manage to detect all the trees correctly and achieve both criteria. In contrast, the GLI and VIgreen segmentation methods are categorized as FN, as the method missed one of the trees in the image.

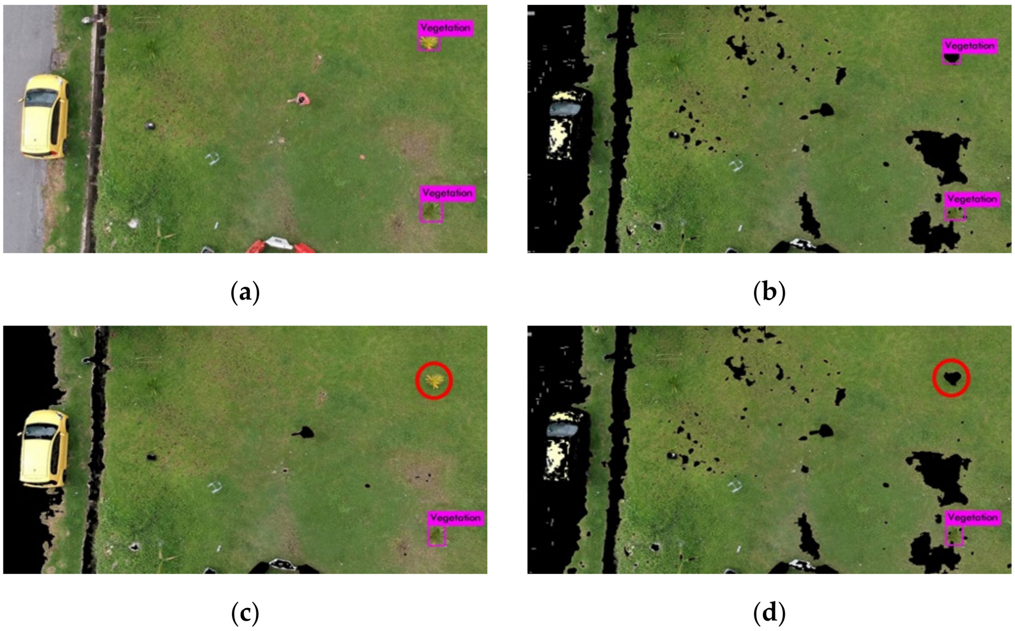
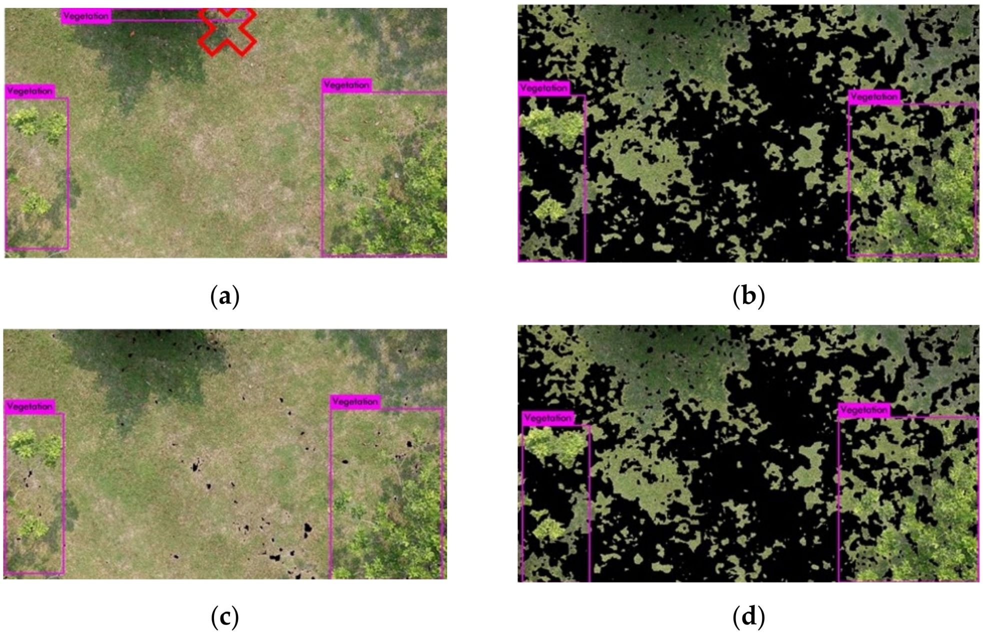


Figure 12. Image from TSI with mostly green fields and a road on the side with a car. (**a**); RGB = TP (all trees correctly detected); (**b**) VARI Seg = TP (all trees correctly detected); (**c**) GLI Seg = FN (1 missing tree); (**d**) VIGREEN Seg = FN (1 missing tree).

The following example was captured on a green field with some portion of tree crowns. There are two different trees located at each of the bottom corners of the image. Figure 13 shows the detection result from the analysis. The results clearly depict that all methods managed to detect both expected trees around the image. However, the RGB method had an additional boundary box located at the top of the image. This particular boundary box only covered the shadow of the tree positioned out of the image. Therefore, in such conditions, the RGB method was limited in differentiating the vegetation shadow on the green land. This limitation might not be a serious problem, depending on the perspective, because an actual tree was located just slightly at the top of the image view. Suppose the method could identify a possible tree through a shadow before the actual tree appears in the image. In that case, it means this method has more sensitivity to possible object arounds and performs pre-awareness checks. Although the RGB method was able to identify both expected trees around the image, the RGB method was evaluated as FN because it did not meet the criterion of having the same amount of the expected boundary box. The three methods were considered TP because they correctly detected all the expected trees and positioned them in the correct area.

Figure 14 mainly consists of a green field with a small part of the tarmac road on the side. In addition, there are non-vegetation construction equipment in the middle of the area. However, our focus is on the two groups of trees located at the right in the image. This expected boundary box has two boxes that cover the large group of trees located on the right and a partially crowned tree situated at the top right. Similar to the last result, all three methods correctly locate both groups of trees and meet the criteria. However, the RGB method is unable to detect the tree located on the top right. The machine learning system could have been confused due to overlapping by the large group of trees and the colour of the background being similar to the tree's leaves. Identifying this is hard because the boundary box for detecting the large tree group is slightly extended beyond the treetop.

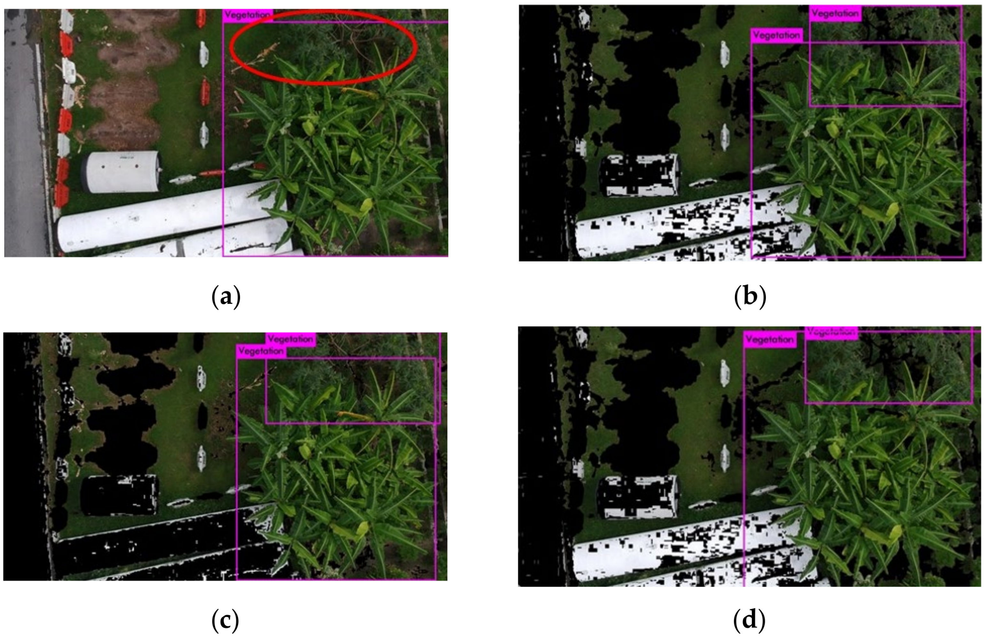
One of the advantages is the segmentation in these examples—the GLI segmentation method is able to segment most of the construction objects positioned in the middle and top left. At the same time, the VIgreen and VARI segmentation methods can only eliminate some portions of the objects. This is a good outcome because those three methods could eliminate non-vegetation objects from the image. It is also suitable for the energy company to perform vegetation encroachment inspection because most of the transmission line in Malaysia is painted in a colour similar to construction objects.



(**a**) (**b**)

(**c**) (**d**)

Figure 13. Image from UNITEN with mostly green fields and some trees. (**a**); RGB = FN (detected 1 tree shadow); (**b**) VARI Seg = TP (all trees correctly detected); (**c**) GLI Seg = TP(all trees correctly detected); (**d**) VIGREEN Seg = TP(all trees correctly detected).



(**a**) (**b**)

(**c**) (**d**)

Figure 14. Images from TSI with mostly green fields and tarmac on the side. (**a**); RGB = FN (1 tree missing); (**b**) VARI Seg = TP (all trees correctly detected); (**c**) GLI Seg = TP(all trees correctly detected); (**d**) VIGREEN Seg = TP(all trees correctly detected).

The confusion matrix was again constructed on a total of 3360 test images, whereby each of the methods was evaluated similarly to the example above. The displayed result in Table 5 showed that the original RGB method achieved at least 83.6% accuracy with a 16.3% error rate. Therefore, the VARI segmentation method was the only one that successfully achieved a similar and slightly higher accuracy rate from the original RGB method. It correctly predicted closer to an equal amount of trees in the images and showed no vegetation images better than the original RGB method.

Table 5. Confusion matrix result.

n = 3360		RGB Method		*n* = 3360		GLI + YOLOv3 (Segmentation)	
		No	**Yes**			**No**	**Yes**
Actual	No	890		Actual	No	929	
	Yes	550	1920		Yes	605	1826
Accuracy	0.836	Error Rate	0.164	Accuracy	0.82	Error Rate	0.18
***n* = 3360**		**VARI + YOLOv3 (Segmentation)**		***n* = 3360**		**VIgreen + YOLOv3 (Segmentation)**	
		No	**Yes**			**No**	**Yes**
Actual	No	920		Actual	No	861	
	Yes	522	1918		Yes	721	1778
Accuracy	0.845	Error Rate	0.155	Accuracy	0.785	Error Rate	0.215

Regarding accuracy rate, the VARI segmentation method is slightly better than the original RGB method by just 0.01 difference. The RGB method had correctly predicted 1920 images having vegetation, while 890 images correctly predicted no vegetation. However, VARI predicted 1918 images of vegetation, but correctly predicted 920 images without any. This result clearly shows VARI leading with 30 images of correctly predicted no-vegetation, compared to the RGB image. The GLI segmentation method achieved an 82% accuracy rate in this experiment and had better prediction in no-vegetation images, close to the VARI segmentation method. However, incorrect prediction of vegetation for the GLI segmentation was slightly higher than the VARI segmentation with less than 100 images, but the GLI had a better prediction of no tree images than the VARI segmentation. In contrast, the VIgreen segmentation method achieved below 80% accuracy in the experiment. This segmentation method in VIgreen had inaccurate predictions in both no-trees and visible tree images and impacted the higher error rate of the prediction, compared to the other methods in the experiment.

6. Discussion

The experiment was conducted to identify how the proposed methods predicted vegetation with RGB UAV images. The evaluation criteria for the experiment consisted of tree availability and the equal number of boundary boxes that was correctly located within the tree objects. According to the results of the confusion matrix, the RGB method and VARI + YOLOv3 performed well throughout the experiment with above 83% accuracy rate. The additional VI segmentation stage on VARI + YOLOv3 was able to eliminate most non-vegetation objects based on the colour in the image before performing a detection. This was able to reduce the possibility of mis-detecting non-vegetation objects as vegetation.

However, the proposed methods have their own limitations. The limitation of VARI and VIgreen segmentation method is elimination of yellowish trees due to its formula of being less sensitive to that colour, resulting in a negative value. The RGB + YOLOv3 method had difficulty detecting trees with shadows in the background. One of the common limitations for every focused method in the experiment is the inability to fully detect individual trees in crowded vegetation areas.

A few strengths are to be highlighted. Firstly, this experiment has proven that the colour of the object could improve detection accuracy. The assistance of the VI segmentation phase eliminates non-vegetation pixel before performing detection using YOLO. Unfortunately, it is a disadvantage to the VIgreen and VARI segmentation methods because the proposed methods eliminate most yellowish trees. In conclusion, the proposed methods have shown performance with acceptable accuracy. Based on the above observations, the YOLO model had difficulties in detecting vegetation on a grayscale VI image, compared to a segmentation. Although the proposed methods achieved good performance, each of the methods tested showed strengths and weaknesses in certain conditions.

7. Implications

This research aimed to test and find the most suitable combination of low-cost off-the-shelf UAV and image processing methods for automated vegetation detection. This combination was studied to help small industries and start-up projects perform vegetation detection using a computer with a suitable processing method with data captured by a low-cost drone because most drone technologies in Malaysia are costly to invest in. In addition, advanced sensors for drone payload is more expensive to acquire, compared to a normal RGB camera [41]. According to [42], some industry domains has started to adopt off-the-shelf drones to replace conventional methods, such as human patrolling. However, the current pitfalls in providing intelligent data analytics on vegetation inspection need to be tackled first [41,43]. On that account, our research aims to identify a suitable processing method to perform vegetation inspection using RGB images. The combination of the UAV and the proposed processing method opens up a new selection for the start-up industry to perform vegetation inspection tasks effectively.

8. Future Work

From the perspective of our research, an improvement in accuracy should be researched in the future. Further experiments on different combinations could help test the remaining VI and machine learning methods, such as the modified green-red vegetation index and the new green-red vegetation index, as proposed by [44]. Since the experiment found that the proposed methods had difficulty detecting smaller objects, an improvement in the YOLOv3 or an alternative model could be further investigated. Furthermore, modified excess green is another algorithm that can be introduced to differentiate green vegetation from RGB images, but it was excluded from our studies.

9. Conclusions

This experiment showed that the original RGB method and VARI segmentation with YOLOv3 achieved the most accurate results. They were capable of performing vegetation detection by implementing the proposed algorithm methods. The function of the VI segmentation stage was mainly to eliminate more non-vegetation objects from the image based on negative values. The results confirm that the proposed methods have the least misdetection errors in the test images. However, each of the methods tested in this research had strengths and limitations in certain conditions, such as surrounding objects and colour of the objects.

Author Contributions: Conceptualization, W.H. and L.S.E.; methodology and investigation, W.H., L.S.E. and R.I.; resources, S.D. and L.S.E.; writing—original draft preparation, L.S.E., W.H. and G.A.; writing—review and editing, W.H., L.S.E., G.A. and Y.B.; supervision, W.H., R.I. and A.A.A.; project administration, W.H., R.I., A.A.A. and A.H.; funding acquisition, W.H., R.I. and S.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by UNITEN Research and Development (URND), project code U-TS-RD-19-16.

Data Availability Statement: Not applicable.

Acknowledgments: This research was supported by UNITEN Research and Development (URND) grant, project code U-TS-RD-19-16. We acknowledge the use of facilities and equipment provided by Micro Multi Copter Aero Science & Technology and Kembara Impian Technologies Sdn. Bhd. We would like to thank every party stated above for providing countless help and assisting in this research.

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

References

1. Li, Z.; Walker, R.; Hayward, R.; Mejias, L. Advances in Vegetation Management for Power Line Corridor Monitoring Using Aerial Remote Sensing Techniques. In Proceedings of the 2010 1st International Conference on Applied Robotics for the Power Industry, Montreal, QC, Canada, 5–7 October 2010; pp. 1–6.
2. Kalapala, M. Estimation of Tree Count from Satellite Imagery through Mathematical Morphology. *Int. J. Adv. Res. Comput. Sci. Softw. Eng.* **2014**, *4*, 490–495.
3. Berni, J.A.J.; Zarco-Tejada, P.J.; Suárez, L.; González-Dugo, V.; Fereres, E. Remote Sensing of Vegetation From UAV Platforms Using Lightweight Multispectral and Thermal Imaging Sensors. *Int. Arch. Photogramm. Remote Sens. Spat. Inform. Sci.* **2009**, *38*, 6.
4. Feng, Q.; Liu, J.; Gong, J. UAV Remote sensing for urban vegetation mapping using random forest and texture analysis. *Remote Sens.* **2015**, *7*, 1074–1094. [CrossRef]
5. Kaneko, K.; Nohara, S. Review of Effective Vegetation Mapping Using the UAV (Unmanned Aerial Vehicle) Method. *J. Geogr. Inf. Syst.* **2014**, *6*, 733–742. [CrossRef]
6. Watanabe, Y.; Kawahara, Y. UAV Photogrammetry for Monitoring Changes in River Topography and Vegetation. *Procedia Eng.* **2016**, *154*, 317–325. [CrossRef]
7. Xue, J.; Su, B. Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications. *J. Sens.* **2017**, *2017*, 1353691. [CrossRef]
8. Gitelson, A.A.; Kaufman, Y.J.; Stark, R.; Rundquist, D. Novel algorithms for remote estimation of vegetation fraction. *Remote Sens. Environ.* **2002**, *80*, 76–87. [CrossRef]
9. Louhaichi, M.; Borman, M.M.; Johnson, D.E. Spatially Located Platform and Aerial Photography for Documentation of Grazing Impacts on Wheat. *Geocarto Int.* **2001**, *16*, 65–70. [CrossRef]
10. Harbaš, I.; Prentašić, P.; Subašić, M. Detection of roadside vegetation using Fully Convolutional Networks. *Image Vis. Comput.* **2018**, *74*, 1–9. [CrossRef]
11. Li, L. The UAV intelligent inspection of transmission lines. In Proceedings of the 2015 International Conference on Advances in Mechanical Engineering and Industrial Informatics, Zhengzhou, China, 11–12 April 2015.
12. Gopinath, G. Free data and Open Source Concept for Near Real Time Monitoring of Vegetation Health of Northern Kerala, India. *Aquat. Procedia* **2015**, *4*, 1461–1468. [CrossRef]
13. Modzelewska, A.; Stereńczak, K.; Mierczyk, M.; Maciuk, S.; Bałazy, R.; Zawiła-Niedźwiecki, T. Sensitivity of vegetation indices in relation to parameters of Norway spruce stands. *Folia For. Pol. Ser. A* **2017**, *59*, 85–98. [CrossRef]
14. Dustin, M.C. *Monitoring Parks with Inexpensive UAVS: Cost Benefits Analysis for Monitoring and Maintaining Parks Facilities*; University of Southern California: Los Angeles, CA, USA, 2015.
15. Xiao, C.; Qin, R.; Huang, X. Treetop detection using convolutional neural networks trained through automatically generated pseudo labels. *Int. J. Remote Sens.* **2020**, *41*, 3010–3030. [CrossRef]
16. Di Leo, E. Individual Tree Crown detection in UAV remote sensed rainforest RGB images through Mathematical Morphology. *Remote Sens.* **2019**, *11*, 1309.
17. Mckinnon, T.; Hoff, P. Comparing RGB-Based Vegetation Indices with NDVI for Drone Based Agricultural Sensing. *Agribotix. Com.* **2017**, *21*, 1–8.
18. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767.
19. Grivina, Y.; Andri, S.; Abdi, S. Analisis Pengunaan Saluran Visible Untuk Estimasi Kandungan Klorofil Daun Pade Dengan Citra Hymap. (Studi Kasus: Kabupaten Karawang, Jawa Barat). *J. Geod. Undip.* **2016**, *5*, 200–207.
20. Barbosa, B.D.; Ferraz, G.A.; Gonçalves, L.M.; Marin, D.B.; Maciel, D.T.; Ferraz, P.F.; Rossi, G. RGB vegetation indices applied to grass monitoring: A qualitative analysis. *Agron. Res.* **2019**, *17*, 349–357.
21. Tucker, C.J. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.* **1979**, *8*, 127. [CrossRef]
22. Bassine, F.Z.; Errami, A.; Khaldoun, M. Vegetation Recognition Based on UAV Image Color Index. In Proceedings of the 2019 IEEE International Conference on Environment and Electrical Engineering and 2019 IEEE Industrial and Commercial Power Systems Europe, Genova, Italy, 10–14 June 2019; pp. 1–4.
23. Motohka, T.; Nasahara, K.N.; Oguma, H.; Tsuchida, S. Applicability of Green-Red Vegetation Index for Remote Sensing of Vegetation Phenology. *Remote Sens.* **2010**, *2*, 2369–2387. [CrossRef]
24. Sanseechan, P.; Saengprachathanarug, K.; Posom, J.; Wongpichet, S.; Chea, C.; Wongphati, M. Use of vegetation indices in monitoring sugarcane white leaf disease symptoms in sugarcane field using multispectral UAV aerial imagery. *IOP Conf. Ser. Earth Environ. Sci.* **2019**, *301*, 012025. [CrossRef]

25. Mokarram, M.; Hojjati, M.; Roshan, G.; Negahban, S. Modeling The Behavior of Vegetation Indices in the Salt Dome of Korsia in North-East of Darab, Fars, Iran. *Model. Earth Syst. Environ.* **2015**, *1*, 27. [CrossRef]
26. Mokarram, M.; Boloorani, A.D.; Hojati, M. Relationship Between Land Cover And Vegetation Indices. Case Study: Eghlid Plain, Fars Province, Iran. *Eur. J. Geogr.* **2016**, *7*, 48–60.
27. Schneider, P.; Roberts, D.A.; Kyriakidis, P.C. A VARI-based Relative Greenness from MODIS Data for Computing the Fire Potential Index. *Remote Sens. Environ.* **2008**, *112*, 1151–1167. [CrossRef]
28. Ancin-Murguzur, F.J.; Munoz, L.; Monz, C.; Hausner, V.H. Drones as a tool to monitor human impacts and vegetation changes in parks and protected areas. *Remote Sens. Ecol. Conserv.* **2020**, *6*, 105–113. [CrossRef]
29. Larrinaga, A.R.; Brotons, L. Greenness Indices from a Low-Cost UAV Imagery as Tools for Monitoring Post-Fire Forest Recovery. *Drones* **2019**, *3*, 6. [CrossRef]
30. Csillik, O.; Cherbini, J.; Johnson, R.; Lyons, A.; Kelly, M. Identification of Citrus Trees from Unmanned Aerial Vehicle Imagery Using Convolutional Neural Networks. *Drones* **2018**, *2*, 39. [CrossRef]
31. Suarez, P.L.; Sappa, A.D.; Vintimilla, B.X.; Hammoud, R.I. Image vegetation index through a cycle generative adversarial network. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops 2019, Long Beach, CA, USA, 16–17 June 2019.
32. Li, W.; Fu, H.; Yu, L. Deep convolutional neural network based large-scale oil palm tree detection for high-resolution remote sensing images. In Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Fort Worth, TX, USA, 23–28 July 2017; pp. 846–849.
33. Kestur, R.; Angural, A.; Bashir, B.; Omkar, S.N.; Anand, G.; Meenavathi, M.B. Tree Crown Detection, Delineation and Counting in UAV Remote Sensed Images: A Neural Network Based Spectral–Spatial Method. *J. Indian Soc. Remote Sens.* **2018**, *46*, 991–1004. [CrossRef]
34. Weinstein, B.G.; Marconi, S.; Bohlman, S.; Zare, A.; White, E. Individual tree-crown detection in rgb imagery using semi-supervised deep learning neural networks. *Remote Sens.* **2019**, *11*, 1309. [CrossRef]
35. Di Gennaro, S.F.; Toscano, P.; Cinat, P.; Berton, A.; Matese, A. A low-cost and unsupervised image recognition methodology for yield estimation in a vineyard. *Front. Plant Sci.* **2019**, *10*, 559. [CrossRef]
36. Kamilaris, A.; Prenafeta-Boldú, F.X. Deep learning in agriculture: A survey. *Comput. Electron. Agric.* **2017**, *147*, 70–90. [CrossRef]
37. Neupane, B.; Horanont, T.; Hung, N.D. Deep learning based banana plant detection and counting using high-resolution red-green-blue (RGB) images collected from unmanned aerial vehicle (UAV). *PLoS ONE* **2019**, *14*, e0223906. [CrossRef]
38. Der Yang, M.; Tseng, H.H.; Hsu, Y.C.; Tsai, H.P. Semantic segmentation using deep learning with vegetation indices for rice lodging identification in multi-date UAV visible images. *Remote Sens.* **2020**, *12*, 633. [CrossRef]
39. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas, NV, USA, 27–30 June 2016.
40. Bayr, U.; Puschmann, O. Automatic detection of woody vegetation in repeat landscape photographs using a convolutional neural network. *Ecol. Inform.* **2019**, *50*, 220–233. [CrossRef]
41. Haroun, F.M.E.; Deros, S.N.M.; Din, N.M. A review of vegetation encroachment detection in power transmission lines using optical sensing satellite imagery. *Int. J. Adv. Trends Comput. Sci. Eng.* **2020**, *9*, 618–624. [CrossRef]
42. Ab Rahman, A.A.; Jaafar, W.S.; Maulud, K.N.; Noor, N.M.; Mohan, M.; Cardil, A.; Silva, C.A.; Che'Ya, N.N.; Naba, N.I. Applications of Drones in Emerging Economies: A case study of Malaysia. In Proceedings of the 2019 6th International Conference on Space Science and Communication (IconSpace), Johor Bahru, Malaysia, 28–30 July 2019.
43. Noor, N.M.; Abdullah, A.; Hashim, M. Remote sensing UAV/drones and its applications for urban areas: A review. *IOP Conf. Ser. Earth Environ. Sci.* **2018**, *169*, 012003. [CrossRef]
44. Zhang, X.; Zhang, F.; Qi, Y.; Deng, L.; Wang, X.; Yang, S. New research methods for vegetation information extraction based on visible light remote sensing images from an unmanned aerial vehicle (UAV). *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *78*, 215–226. [CrossRef]