International Journal of Information & Computation Technology.
ISSN 0974-2239 Volume 13, Number 1 (2023), pp. 15-25

https://dx.doi.org/10.37624/IJICT/13.1.2023.15-25

Deep Learning Approach for Auto Counting Complex Plants

VijayaKumar Balla and Ramesh Cheripelli

Assistant Professors, Dept of IT
G Narayanamma Institute of Technology and Science, Hyderabad, India
E-mail: vijayballa504@gmail.com, chramesh23@gmail.com

Abstract

In order to address the issues affecting crops and their pro-ductivity, plant phenotyping is one of the emerging research topics that requires attention. In this paper, we use Convolutional Neural Networks architecture to count the plants in agricultural areas. Regression is used in place of classification to estimate the number of plants in a field pho-tograph. This eliminates the requirement to know (or estimate) the max-imum anticipated number of plants. The CNN will be trained on these images and evaluated using them. Our tests demonstrate that using the Inception-v3 CNN architecture, we can get a Mean Error as low as pos-sible.

Keywords: Plant Count, Deep learning, Phenotyping, Neural Net-work, Regression

1 Introduction

Plant density, or the number of plants per square metre, and crop productivity, or grain yield, are closely related in significant seeds like cotton seed [1, 2]. There-fore, efforts to improve crops by genetic and managerial means frequently centre on the plants that are being measured. Usually, hand counting methods are used for this, which calls for greater labour. Such methods are time-consuming, expensive, prone to mistakes, and slow. Thus, there is a definite demand for technology that make it possible to count plants accurately and efficiently at various crop growth stages

Using the Convolutional Neural Networks (CNN) architecture, we solve the phenotyping job of counting crop plants in a field in this research. To evaluate number of plants, we employ multiple CNN architectures with few alterations. The number of plants

There is a respectable corpus of work on counting plants and other items in the literature. By using 3D laser scanning scans from a vehicle that is being driven through the field, plants are counted in [3]. The derived plant skeleton in [4] is used to estimate the population of maize plants and their locations. A poplar plantation's tree crowns are counted in [5] by minimising an energy func-tion that represents the plants as uniformly spaced ellipsoids. Using an estimated item density function, objects are tallied in [6]. The training process needs that the object localisation be provided (i. e., pointing at the object).

In order to learn illustrations of data with various degrees of thought, DL is a class of machine learning methods that employs a stack of many processing layers, with each succeeding layer using output from preceding layer as input. DL models are frequently created using multilayer neural networks, in which neurons serving as the connections between two successive layers of features basically reflect different parametrized nonlinear transformations. Weights and biases of the neurons, which are multiplied and added to the input, are two examples of model parameters. Up until the target layer or the decision layer, input data is converted forward in a layer-by-layer method. For instance, in order to alter an input image for a challenge of image classification, hierarchical nonlinear transformations and finally image class becomes the output at the target layer.

An primary attribute set for deep neural network (typically picked at ran-dom) is used to start training such a model (DNN). At the target layer, errors are calculated between desired outputs provided by training data labels and the actual outputs for a huge number of instances. Settings are then updated using a feedback methodlogy layer by layer (from target to input) until the target layer reaches a suitable degree of judgement accuracy. For this training phase, error back-propagation algorithm is typically used.

The typical features that a DNN learns have a hierarchical structure, mean-ing that while the lower layers of abstraction capture basic low-complexity fea-tures, the higher layers of abstraction form more complex features by combining the low-complexity features in intricate ways. DNNs have been trained using stochastic gradient descent (SGD) and its derivatives, including minibatch gra-dient descent [8], ADAM [9], and ADMM [10].

Many computer vision tasks have been addressed using deep learning tech-niques [7, 8, 9]. To develop a classifier or other predictions, deep learning uses nonlinear attributes which will themselves trained to describe the input data [8, 10]. Deep learning has recently been applied to distant sensing tasks like target detection, image restoration, and pixel-based classification [11]. There has been research on counting the number of objects in photos using CNNs. In [12], a neural network with two convolutional and two linear layers is utilised to count objects. The MNIST dataset [13] and the UCSD dataset [14] are specifically used for counting digits and pedestrians, respectively. [15] describes two methods for counting items in photos. A density function is clustered using one. The second method makes use of Support Vector Machines (SVM) for the VGG19 CNN's first two fully connected layers and final pooling layer in the CNN [16]. A ResNet architecture variant is suggested in [17] for enumeration and localising automo-biles from satellite photos. [21] uses a custom CNN to degenerate population in compact crowds.

2 Collection of Data

Due to the dearth of publicly accessible datasets for task of plant counting, we created our own dataset for training and testing. This dataset comprises of 4,316 RGB photos of cotton field regions taken by a UAV when it was flying at a 45-meter altitude. Three samples of the RGB images from our dataset are shown in Figure 2.1 which are taken on from July 13, 2022. Each photograph has a resolution of 540 by 140 pixels and between 0 and 20 plants. These pictures were created by cropping orthoimages of a cotton field that were taken by a UAV at different times during the growing season [19]. The UAV cameras are calibrated to establish their internal settings before producing the orthoimages. Triangulation is used to estimate the coordinates of sparse points as well as the location and positioning of each UAV image. [19] describes process used to create the orthoimages. Produced from UAV pictures taken on July 11, 2021.

A panel is one of the field's three primary divisions. Because it is the only part of the central panel that can be groundtruthed, only the upper half is used for analysis. The plants in other parts of the field are too closely packed together to be physically counted. Ranges are divisions within a panel that run parallel to the direction of the plant alignment, while rows are divisions that run perpendicular to that direction.

A part of the orthoimage in Figure 3 with two ranges and five rows, gen-erating ten sub-rows, is an example of a panel with M rows and N ranges and MN sub-rows, which is an aligned ensemble of a few plants. The orthophoto is used to create the images in our dataset in the manner described below. The orthophoto's Region Of Interest (ROI) is first cropped using the user's specified coordinates. The central panel of Figure 2 is ROI for our dataset. The panel & a few pixels of the neighboring soil must be totally included in the ROI, although the coordinates do not need to be extremely exact. Next, we create a mask utilising the segmentation method, we described in [19]. Summation of P (x)=I (x,y) (Equation 2) where each column x of the ROI receives a vertical sum for each. We refer to the "horizontal profile" as ph. An illustration of a horizontal profile is shown in Figure 4. The vertical lines separating the field's ranges are shown by the valleys in ph (). These lines should ideally only cross soil that is not covered by plants and not interfere with any vegetation. In actuality, the rows of field are not perfectly straight, and there may even be unanticipated weeds, so pH at the troughs is higher than zero. The increase around x=22, 745 can be attributed to weeds & grass from external field encroaching on it. We predict the range-separating lines will be located as follows: Xn=X0 + n X (Equation 3) where N is the field's total number of ranges. Due to the low dimensionality of the issue, brute force is used to perform the optimization in Equation (3). Equation (3) makes the assumption that the field's ranges are separated by a constant distance (X). For orthophotos that are well-aligned, this assumption is plausible. The procedure described above is applied to each range independently after the field has been partitioned into ranges. Getting lines that separate each array into rows is the aim here. As we move forward, we only choose the area of mask that corresponds to each array a I (x,y)=1 if I (x,y)=1 (Equation 4)=0 else for all arrays a=0,..., N 2. For each array a, vertical outline is attained p (y)=I (x,y) (Equation 5) where the ROI's picture rows y each have a horizontal sum. We calculate the field row dividing lines in each range r as Y r=Y r + m Y r (Equation 6) m is number of field rows. A grid of NM coordi-nates is produced by the points where the vertical and

horizontal lines connect. The bounding boxes are created with the help of coordinates across sub-row of the plant There is a chance that these bounding boxes will differ in dimensions. All bounding boxes are reformed to the intermediate bounding box magnitude while maintaining identical centroid for each bounding box in order to produce a collection of photos that are the same magnitude. The images in our collection are made up of the pixels inside these bounding boxes.
We gathered photos on July 13, 2022, and July 21, 2022, which are two separate days. The photos from July 13, 2022 were obtained using the bounding box coordinates for July 11, 2021. By physically counting number of plants in each shot, we groundtruthed the dataset. The spreading of ground truth for each time is depicted in Figure 2.3. The other 1,240 photos are for July 11, 2021, while 1,240 are for July 3, 2021. Between these two periods, there is a marginal rise in the number of plants.

Figure 1: Three samples of the RGB images 16 plants each with one sub-row of plants

Figure 2: Cotton Crop field acquired on July 21,2022

Figure 3: Acquired Segment of the orthophoto on July 21, 2022

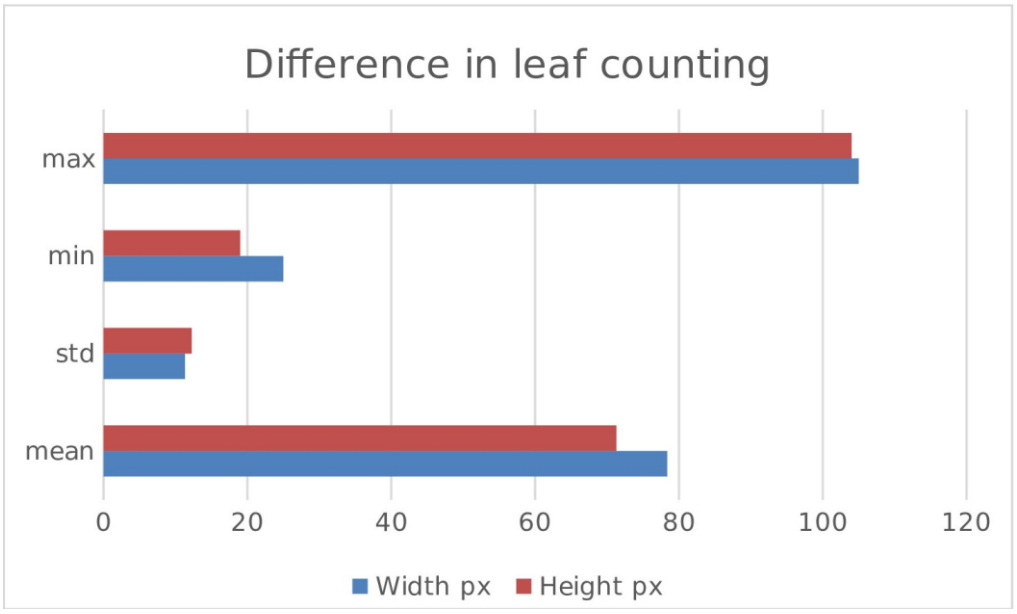

Figure 4: Difference in leaf counting

3 Methodology

3.1 Image Phenotyping System

While gathering data at the plant level yields accurate information correspon-dence, the expert cess is time-consuming, demanding, and occasionally challeng-ing. Crops are densely cultivated in an outside climate. Accessing plants can be difficult and costly. As a result, high throughput plant phenotyping benefits less from information gathering per plant. For instance, UAVs have made it possible to secure aggregate information during current imaging stages. These steps can capture a field's general information in less than an hour, but developing precise information correspondence is more difficult. The physical organisation of the captured images and their naming are time-consuming tasks. The ghostly data collected from UAV symbolisms is projected to a level surface with geocoordi-nates using orthorectification techniques as shown in [4]. We may assign each region with its pixel-level sensor information, which is used for quality assess-ment, using its geocoordinates. We present an RGB image based picture based plant phenotyping system that compiles the information correspondences and evaluates attributes at the plot or plant level. Additionally, this framework in-cludes a web interface and an adequate processing stage, making it simple for plant researchers to access and employ our apparatuses. Figure 3.1 depicts the Linear and Non-Linear Functions of Plant Pixels data generated by the UAV platform as input, and outputs the estimated properties for selected regions. We'll probably add an image showing the number of plants. Our method dif-fers from in that we don't take the plant into account when determining the class number. Last layer of a CNN typically has the same number of neurons as greatest number of plants in image. This a strategy solves a categorization issue. The cross entropy is a typical loss function in a classification task. [9, 18, 23] H (q,p)=p (x) logq (x) (Equation 7) where p (x) is the label and q (x) represents the activations at each of the final layer's x neurons. Using one-hot encoding to set p (x), i.e. p (x)=1, x=ci (Equation 8) the loss function only considers one neuron where Ci is plant count label of ith picture (Ci-th neuron) H (q, p)=log q (Ci) (Equation 9). This suggests that we would ignore any persistent activation while training. P (x) can be understood as the degree of assurance that image goes to class x. When classes are independent, equations (8) and (9) apply. The modules in our dataset are not entirely autonomous, though. The illustration would be a picture with two plants in it. The difficulty to network should be significantly less if the evaluation is three than if it is five. Due to significant label noise in our dataset, this is even worse. When plants are grouped together, the label may differ significantly from the actual value by more than one count.

The need to predetermine the maximum number of classes is another issue with the classification approach. We use the Lp norm as the loss function to be minimised in order to solve these issues: Lp (x, Ci)=|x Ci|p (Equation 10) where x is the projected number of plants. The Mean Squared Error (MSE), a standard loss function for regression, replaces the Lp loss at p=2. This inaccuracy can be seen as the usual mistake we make when estimating the plants count in an image even though p=1.

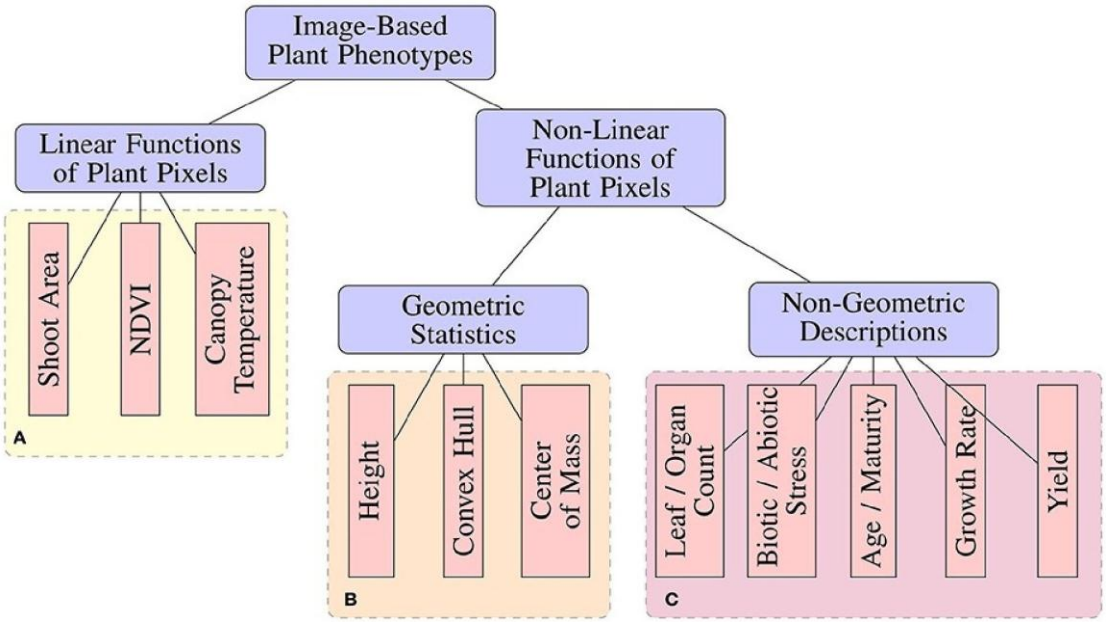

Figure 5: Block Diagram of Image Phenotyping System

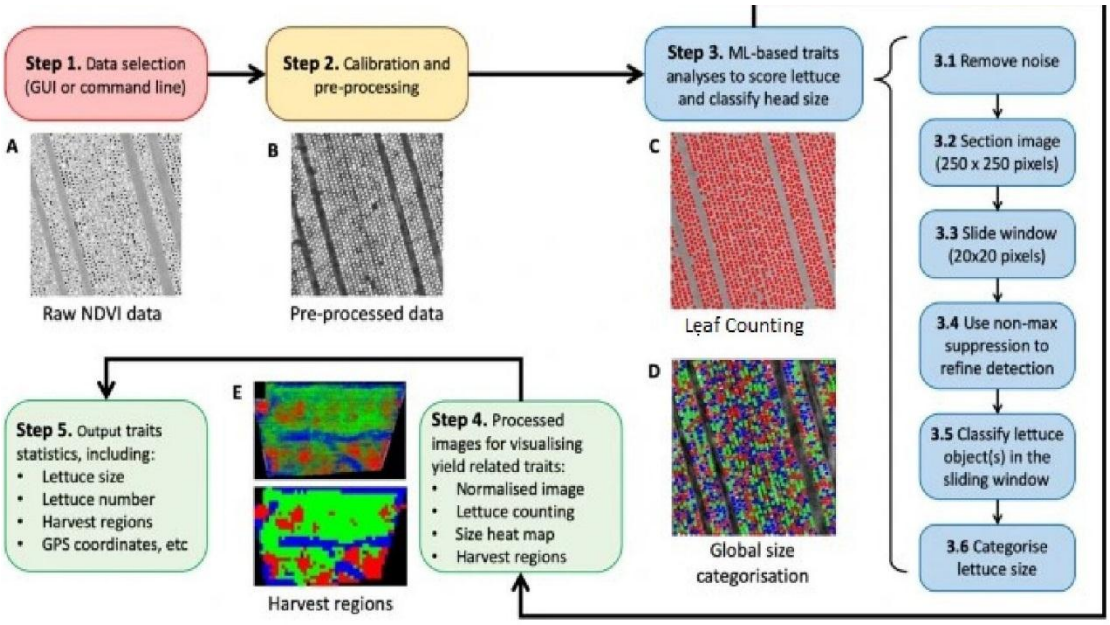

Figure 6: A high-level analysis workflow of leaf counting

4 Experimental Results

Here, we'll discuss the findings of various counting task methodologies and con-trast them with one another. Our comparison will centre on how well various models did when counting on raw image, using the ground truth palms count of 15947. The amount

of the training dataset required to reach the desired accuracy level is another crucial factor to consider

4.1 Regression with density maps

Heatmap on 4.1 was produced as a result of applying the density map regression model to the entire image. The majority of palms had white circles on them that deep neural networks were trained to identify and create from obscured photos. After summarising the resulting heatmap, we were able to count 15784 palms with a 98.9% accuracy rate. Finding coordinates Blob detectors may be used to determine the coor-dinates of the palm centre since the mask we acquire has circles of the same hue that are easily distinguished from one another. We utilised "SimpleBlobDetec-tor" from Bradski, 2000 to get the coordinates. The outcome is displayed on 4.2

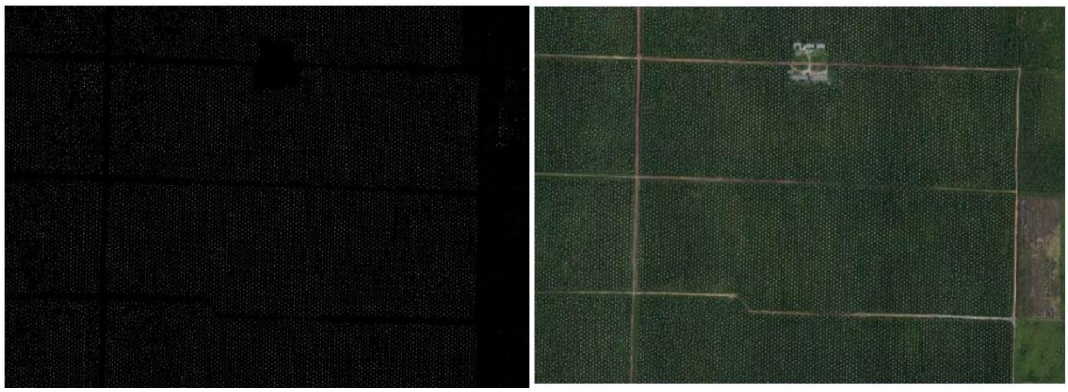

Figure 7: (a)Cotton HeatMap (b) stacked on top of input image

4.1 Regression using a classifier

When the procedure from section 4.4 was applied to the entire dataset, the result was 20448. This gives us a 28% inaccuracy when compared to the 15947 plant instances that make up the ground truth value. This margin is unsatisfactory for the task of counting crops since it introduces a lot of uncertainty when employed in additional plant lifetime modelling and revenue projections. The method might be enhanced by expanding the training set with more data, but this task was excluded from the current study's scope because one of the work's prerequisites was a small training dataset

4.2 Extracting the Foreground

A mask with a surface area of 76023641 square pixels was produced by the foreground extraction technique used in the sugar beet example. We can estimate that there are 37710 plant instances in the field based on the average plant size of 42 48 pixels. With the aid of agricultural professionals, further projections and estimates might be made using this method's 93% accuracy.

4.3 Slider window

Since this strategy was causing several classifier activations on the palm class for a

single plant instance, as can be seen on 4.3, evaluation of the approach when we tried to classify patches was not finished.

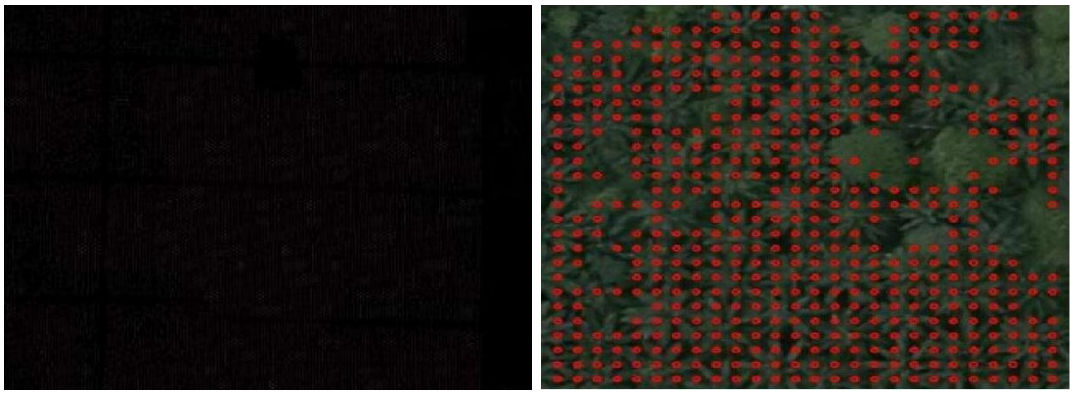

Figure 8: (a)Blob detector results (b) Sliding window iteration result

5 Conclusion

Employing our own collection of RGB UAV images of cotton plants, we demon-strated a method using CNNs to estimate the number of harvest plants in a field. We evaluated how well various neural networks presented themselves for the task of plant counting while making minor design changes. On the testing set, the Inception-v3 design had the smallest Mean Error. Because it produced the least amount of error in our tests, p=1 was the reward for p in the Lp stan-dard. Future work will involve adding more data, using other methodologies for data augmentation, and calculating other plant characteristics like the number of leaves.

References

[1] Dale JL. Banana bunchy top: An economically important tropical plant virus dis-ease. Advances in virus research. 1987; 33:301326.

[2] STEWART L, CAMPAGNOLO D, DANIELLS J, LEMIN C, GOEBEL R, PINESE B, et al. Tropical banana information kit. Nambour: Queensland Department of Primary Industries. 1998.

[3] Bastiaanssen WG, Molden DJ, Makin IW. Remote sensing for irrigated agriculture: examples from research and possible applications. 2000; 46 (2):137155.

[4] Mulla DJ. Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. 2013; 114 (4):358371.

[5] Zhang C, Kovacs JM. The application of small unmanned aerial systems for pre-cision agriculture: a review. Precision agriculture. 2012; 13 (6):693712.

[6] Turner D. Banana plant growth. 1. Gross morphology. Australian Journal of Experimental Agriculture. 1972; 12 (55):209215. https://doi.org/10.1071/EA9720209

[7] Ke Y, Quackenbush LJ. A review of methods for automatic individual tree-crown detection and delineation from passive remote sensing. International Journal of Remote Sensing. 2011; 32 (17):47254747.
[8] Singh K, Sohlberg S, Sokolov V, et al. Conceptual framework for the selection of appropriate remote sensing techniques. 1986.
[9] Pinz A. Tree isolation and species classification. Automated Interpretation of High
[10] Spatial Resolution Digital Imagery for Forestry, Victoria, BC; 1998. p. 127139.
[11] Gougeon FA, Leckie DG, et al. Forest regeneration: Individual tree crown detection techniques for density and stocking assessment. BC; 1998. p. 1012.
[12] Brandtberg T. Automatic individual tree based analysis of high spatial resolution
[13] aerial images on naturally forests. 1999; 29 (10):14641478.
[14] Pollock R. Individual tree In: Proc. of the International Forum on Automated Interpretation of High Spatial Resolution Digital Imagery for Forestry. Victoria, British Columbia, Canada; 1998. p. 2534
[15] Johansen K, Sohlbach M, Sullivan B, Stringer S, Peasley D, Phinn S. Mapping
[16] banana plants from high spatial resolution orthophotos to facilitate plant health assessment. Remote Sensing. 2014; 6 (9):8261 8286.
[17] She T, Ehsani R, Robbins J, Leiva JN, Owen J. Applications of small UAV systems
[18] for tree and nursery inventory management. In: Proceedings of the 12th International Conference on Precision Agriculture, Sacramento, CA, USA; 2014. p. 2023.
[19] Guldogan O, Rotola-Pukkila J, Balasundaram U, Le TH, et al. Automated tree
[20] detection and density calculation using unmanned aerial vehicles. In: 2016 Visual Communications and Image Processing (VCIP). IEEE; 2016. p. 14.
[21] Gnaĺdinger F, Schmidhalter U. Digital counts of maize plants by unmanned aerial
[22] vehicles (UAVs). Remote sensing. 2017; 9 (6):544.
[23] Weinacker H, Koch B, Weinacker R. TREESVIS: A software system for simulta-neous ED-real-time visualisation of DTM, DSM, laser raw data. Remote Sensing and Spatial Information Sciences. 2004; 36:9095.
[24] C. Ramesh, K. V. G. Rao and D. Vasumathi, "Evaluation of Key Management
[25] Scheme Based on Identity," 2016 IEEE International Conference on Advanced Computing, 2016, pp. 545-550, doi: 10.1109/IACC.2016.107. Feburary 27-28,2016
[26] Wallace L, Lucieer A, Watson CS. Evaluating tree detection and segmentation
[27] routines on very high resolution UAV LiDAR data. 2014; 52 (12):76197628.
[28] Mohan M, Silva C, Klauberg C, Jat P, Catts G, Cardil A, et al. Individual tree detection from unmanned aerial vehicle derived canopy height model in an open canopy mixed conifer forest. Forests. 2017; 8 (9):340.
[29] Srestasathiern P, Rakwatin P. Oil palm tree detection with high resolution multi-
[30] spectral satellite imagery. Remote Sensing. 2014; 6 (10):97499774.
[31] Ramesh Cheripelli and A. N. K. Prasannanjaneyulu Generating Automatic

Ground Truth by Integrating Various Saliency Techniques, Pro-ceedings of Second Inter-national Conference on Advances in Computer Engineering and Communication Systems,2022, Springer Singapore, 371384, DOI:10.1007/978-981-16-7389-4-35

[32] Lowe DG. Distinctive image features from scale-invariant keypoints. International journal of computer vision. 2004; 60 (2):91110.

[33] Huang GB, Zhu QY, Siew CK. Extreme learning machine: theory and applications. Neurocomputing. 2006; 70 (1-3):489501.

[34] C. Ramesh; K. Venu Gopal Rao; D. Vasumathi Comparative analysis of applica-

[35] tions of identity-based cryptosystem in IoT Electronic Govern-ment, an International Journal, 2017 Vol.13 No.4, pp.314-323 DOI: 10.1504/EG.2017.088003

[36] LeCun Y, Bengio Y, et al. Convolutional networks for images, speech, and time

[37] series. The handbook of brain theory and neural networks. 1995; 3361 (10):1995.

[38] LeCun Y, Bengio Y, Hinton G. Deep learning. nature. 2015; 521 (7553):436.

[39] Pan SJ, Yang Q, et al. A survey on transfer learning. IEEE Transactions on knowl-edge and data engineering. 2010; 22 (10):13451359.