Semantic segmentation of in-field cotton bolls from the sky using deep convolutional neural networks

Article · March 2022

DOI: 10.1016/j.atech.2022.100045

Naseeb Singh
Indian Institute of Technology Kharagpur

V.K. Tewari
Indian Institute of Technology Kharagpur

Laxmi Kant Dhruw
Indian Institute of Technology Kharagpur

Huidrom Dayananda Singh
Indian Council of Agricultural Research

Smart Agricultural Technology 2 (2022) 100045

Contents lists available at ScienceDirect

Smart Agricultural Technology

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

Semantic segmentation of in-field cotton bolls from the sky using deep convolutional neural networks

Naseeb Singh [a,c,*], V.K. Tewari [a], P.K. Biswas [b], L.K. Dhruw [a], C.M. Pareek [a], H. Dayananda Singh [c]

[a] Department of Agricultural and Food Engineering, IIT Kharagpur, Kharagpur 721302, India
[b] Department of Electronics and Electrical Communication Engineering, IIT Kharagpur, Kharagpur 721302, India
[c] Scientist, Division of System Research and Engineering, ICAR RC for NEH Region, Umiam, Meghalaya 793103, India

ARTICLE INFO

Keywords:
Cotton boll recognition
Sky detection
Semantic segmentation
Convolutional neural network
Robotic harvesting
Image segmentation

ABSTRACT

Manually picking of cotton bolls is a tedious, costly, and labor-intensive task, while harvesting using machines results in higher harvesting losses. By keeping selective picking in mind to maintain fiber quality, minimize harvesting losses, and tackle the shortage of farm labor in near future, cotton harvesting robots seem to be a better alternative in coming years in both developing and developed countries. For the cotton harvesting robot, cotton boll recognition with minimum errors is a foremost and challenging task. While recognizing cotton bolls, false-positive errors occur due to sky interference. In present study, convolutional neural networks were used to segment and discriminate the cotton bolls pixels from sky pixels. For that, three fully convolutional neural networks namely VGG16, InceptionV3, and ResNet34 were used as encoders and trained. These trained neural networks models were evaluated using the intersection-over-union (IoU), F1-score, precision, and recall metrics. All proposed models were tested on a cotton-sky dataset and achieved an IoU score of above 81% and 80% for cotton bolls and sky, respectively. InceptionV3 model outperforms with an IoU score of 84.5% and 80.67% for cotton bolls and sky, respectively with a segmentation time of 1.07 s. For the cotton dataset, proposed models achieved an IoU score of above 90% for cotton bolls and the InceptionV3 model outperforms with an IoU score of 93.29%. It can be concluded that the InceptionV3 model segmented cotton bolls and sky with higher accuracy, and low error rates and, hence can be deployed to cotton harvesting robots effectively.

1. Introduction

Cotton (*Gossypium hirsutum* L.*)* is an important cash crop for farmers and at present can be harvested mechanically or manually. In developed countries like the USA, Australia, etc. fully mechanical harvesting of cotton is done while in some countries like Uzbekistan, 30-40% of cotton is picked using the machine [9]. In developing countries like India, as the mechanization level in cotton harvesting is nil [43], fully manual picking of cotton is practiced. Though manual picking of cotton preserves better fiber characteristics compared to machine picking, it is a time-consuming and labor-intensive task. As predicted by Mehta et al. [43], the percentage of agricultural workers to total workers in India will be decreased to 40.6% by the year 2020, due to which the availability of agricultural workers will decrease, while labor costs will increase. Timeliness in cotton harvesting is important to avoid losses and maintain the quality of cotton fiber [20] which will be adversely affected by the non-availability of laborers at the desired time. Another aspect of manual picking is that the cotton pickers are exposed to various hazards during picking like pesticide residue, fine dust particles, etc. and later faces health problems like headache, stomach pain, vomiting, asthma, etc [8,32]. In the case of harvesting using machines, artificial defoliation of cotton plants should be carried out before harvesting [60] which adds expenses to the cotton cultivators. Additionally, after the application of defoliation chemicals, farmers need to wait a few additional days to harvest cotton bolls, and this waiting time exposes the open bolls to harsh climatic conditions that may lead to losses in terms of quality and/or quantity of cotton. So, manual cotton harvesting is a tedious, costly, labor-intensive task and may cause health hazards, while harvesting using machines are expensive as well as inorganic because of the use of defoliation chemicals and may lead to losses because of harsh climatic conditions occasionally. By keeping selective picking in mind to maintain fiber quality as well as to replace human pickers, harvesting robots can be a good alternative for cotton harvesting.

In the past, several researchers have studied fruit harvesting using robots [12,14,18,36,40,52,56,72,73,81,85] in which their first major task was to recognize the fruit for harvesting using machine vision. In agriculture, various image processing techniques were used by numerous researchers [13,17,24,26,28,42,50,53,54,58,59,63,70,75] to

* Corresponding author at: Department of Agricultural and Food Engineering, IIT Kharagpur, Kharagpur 721302, India.
E-mail address: naseeb501@gmail.com (N. Singh).

https://doi.org/10.1016/j.atech.2022.100045
Received 19 September 2021; Received in revised form 9 March 2022; Accepted 11 March 2022

segment the region of interest from the complex background for which color, shape and texture features were extracted using hand engineered-based method and used as an individual basis or as a combination of multiple features.

For a cotton harvesting robot, the first foremost task will be the segmentation of cotton bolls in the field [13,44] which is a challenging task as the shape and size of cotton bolls are non-uniform [37]. Also, due to varying illumination conditions in the field, cotton bolls appear bright and dull during high and low illumination, respectively [37] which may cause errors in cotton recognition. In recent times, several image processing approaches were used for the recognition of cotton bolls [19,29,58,62,77,83] for which cotton recognition features were selected manually and, in some cases, required complex pre-processing of raw images. Features extracted manually are often low-dimensional [76] and important high-dimensional features that will be useful to discriminate cotton bolls from complex background are difficult to extracted effectively. In recent years, convolutional neural networks (CNNs) have demonstrated remarkable performance in segmentation of region of interest from the images having complex backgrounds by extracting high-dimensional features through self-learned features. Conceptually, CNNs contain a hierarchy of self-learned features, all of which are based on less abstract features from the previous layers of the network and achieved impressive results in image segmentation in agriculture. For semantic segmentation, fully convolutional neural networks are utilized which carried out classification at pixel level by accepting raw image as input and provide the corresponding predicted segmented image as output. Numerous researchers achieved promising results in semantic segmentation using fully convolutional neural networks [11,15,68,71,82]. Kestur et al. [31], developed a deep sematic segmentation architecture based on a fully convolutional neural network to detect as well as count the mangoes and achieved an F1-score of 66%. Majeed et al. [41], used a convolutional neural network for semantic segmentation of apple tree trunk and branch and achieved mean intersection-over-union (mIoU) scores of 59% and 44%, respectively. Azizi et al. [6], performed a semantic segmentation task on soil clods for identification purposes in which they used the VGG16 CNN model as an encoder and achieved a mean intersection-over-union (mIoU) score of 80.50%. From the above works, it can be concluded that image segmentation based on CNNs can overcome the constraints of conventional image segmentation methods.

Hence, this study aimed to use the deep convolutional neural networks for semantic segmentation of in-field cotton bolls from the images having sky in the background by extracting high-dimensional features automatically using self-learning features which else be a difficult task if performed using present conventional image segmentation methods. As the required output is semantic segmentation, therefore, the fully convolutional neural network was used as encoder as well as decoder which was trained through transfer learning. The performance of trained convolutional neural network (CNN) models was evaluated in terms of intersection over union, F1-score, precision, and recall values. The performance of trained CNN models was also compared with existing state-of-the-art methods for cotton recognition.

The rest of the paper is organized as follows. In Section 2, the work related to cotton bolls recognition is presented, in Section 3, image dataset, CNN models architecture, training of neural networks, and metrics used for performance evaluation are presented. The results obtained by the trained CNN models to segment cotton bolls and the sky as well as limitations of developed CNN models are presented in Section 4. In the end, Section 5 concludes the findings of present study.

2. Related work

Out of color, texture, and shape features used in conventional image recognition methods for the agriculture-related region of interest segmentation tasks, color is the salient feature of cotton bolls and used by several researchers for cotton bolls segmentation in the past. Jin-Shuai et al. [29], used YCbCr color space and fisher discrimination analysis to segment cotton bolls from the background whereas Wang et al. [77] used the color subtraction method to detect cotton bolls and freeman chain coding to remove noises. Yeom et al. [83] proposed an open cotton boll detection algorithm in which cotton boll was identified based on spatial features of each region growing segment and derived threshold values were used for binary cotton boll classification. Singh et al. [58] developed four image processing algorithms based on color features to segment the cotton bolls from the background under varying illumination conditions. Sun et al. [62] proposed image processing algorithms in which double-thresholding with region growth algorithm combining color and spatial features was applied to segment cotton bolls. Feng et al. [19] developed a model to estimate the cotton crop yield using UAV-based multi-sensor imagery for which eight image features were extracted. Singh et al. [58] in their study observed that one of the major challenges in the correct identification of cotton bolls in the field is the errors due to sky interference in the scene which may lead to false-positive detection of the sky as cotton bolls. The image dataset of above mentioned studies [19,29,58,62,83] did not contain the sky in background, and hence, developed algorithms were not evaluated for false-positive errors which can be originated due to sky interference during cotton boll segmentation. From Fig. 1, it can be observed that for RGB, HSV, or YCbCr color space model, the intensity values of color components of cotton bolls are in the same range as of their respective color components of the sky due to which segmentation of cotton bolls using color features will lead to the multiple numbers of false-positive detections and hence, cotton bolls and sky are indistinguishable using color features solely. Therefore, in addition to the color feature, other high-dimensional features should be extracted which is a difficult task using present conventional image segmentation methods. An alternative is the use of convolutional neural networks which can extract high-dimensional features through self-learned features. Li et al. [38] used a deep fully convolutional network for in-field cotton segmentation in which they achieved an intersection-over-union value of 70.4% and 59.1% for sunny conditions and highlight conditions, respectively. Li et al. [37] employed a region-based semantic segmentation method for in-field cotton detection in which they achieved an intersection-over-union value of 73.5% for cotton bolls. Tedesco-Oliveira et al. [69] developed an automatic deep learning system to predict cotton yield and observed accuracy of above 80%. But these studies did not evaluate the discrimination accuracy of models between cotton bolls and sky. Hence, a study is needed for semantic segmentation of in-field cotton bolls from the images having sky in the background.

3. Materials and methods

3.1. Experimental site and image acquisition system

Haryana is a major cotton-growing state of Northern India [55], and the cotton images analyzed in the present study were captured from a cotton field of *Kharanti* village (29°01’30”N, 76°27’12”E), Rohtak district, India. Cotton seeds were manually sown while maintaining the row to row and plant to plant spacing of 0.67 and 0.30 m, respectively, which results in approximately 49,300 plants per hectare. The intercropping practice was not implemented in the field. In 170-175 days, cotton plants reached their maturity stage, and the height of these plants at this stage was around 1.50 m. A total of one hundred numbers of RGB color images (640 × 480 pixels resolution) were captured having the sky in the background. For acquiring images, a digital camera (Logitech Webcam C270, 0.9 megapixels, CMOS as image sensor type, fixed focal length of 4.6 mm, diagonal field of view of 55°), and a laptop (8 GB RAM, Intel Core i5 CPU, and Windows 10 operating system) were used. While capturing the images, the height of the camera from the ground was kept between 0.85 m to 1.35 m while the distance between the camera and capture target varies between 0.3 to 0.65 m. The images were captured in the 2nd week of October 2020 during the daytime from 8:00 to 17:00 h (average illumination of 55,029 lx) between October 9th and

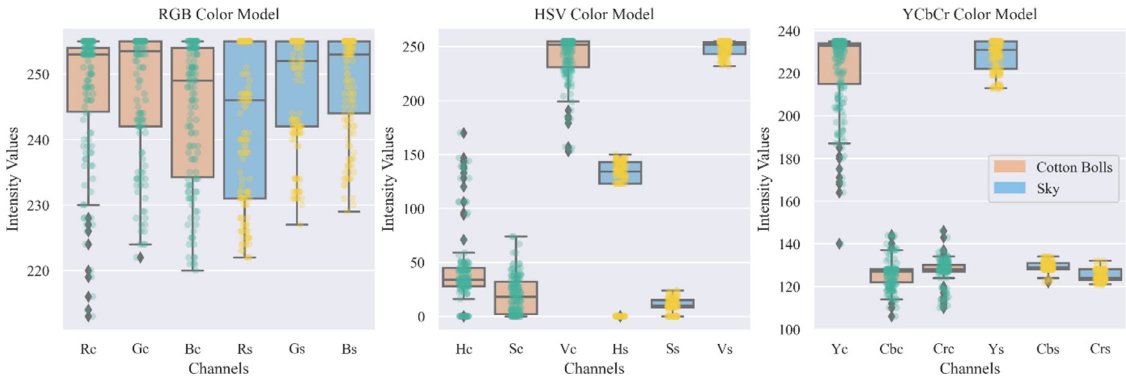

Fig. 1. Color channels of cotton bolls, and sky in RGB, HSV, and YCbCr color space models.

October 12th. During this time, the cotton bolls of the experimental field were due for harvesting in a week.

3.2. Dataset construction and augmentation

For the semantic segmentation task using convolutional neural networks, two sets of images are required. One set contains the RGB images used as input to CNNs and the other set contains the corresponding labeled images. In labeled images, each pixel is assigned a value corresponding to the classification class and this image is called ground truth image. In this study, an annotation app called '*apeer*' provided by *ZEISS Microscopy* [5] was used to label each pixel of an image with one of three classes: 'Background', 'Cotton', and 'Sky'. The background class consists of leaves, stems, soil, plastics, etc. These labeled images have a direct and significant influence on CNNs accuracy and hence labeling was done with high accuracy for this study which, therefore, consumes a lot of resources in terms of time, human power, etc. As the training of CNNs needs a large number of labeled images [86], but due to practical limitations of resources in data collection and providing the corresponding labeled images, image augmentation technique was implemented in this study. For this purpose, rotation (-30° to 30°), horizontal shifting (64 pixels), vertical shifting (48 pixels), horizontal flip, and vertical flip of images with random values were carried out. In past, numerous researchers [6,14,15,86] used this technique for image augmentation and achieved impressive accuracy for semantic segmentation. The dataset consists of a total of 800 images (using a single raw image seven augmented images were produced) which contain cotton bolls and sky Fig. 2. shows the sample of augmented raw images and corresponding labeled images used for CNNs training in the present study.

3.3. Convolutional encoder-decoder network

Generally, for semantic segmentation, a convolutional encoder-decoder network is used [7] in which the encoder computes higher-level features as the number of receptive fields increases with the network depth and the decoder compute feature maps of progressively increasing resolution *via* up-sampling. Skip connections [51], recurrent connections [48], and larger convolutional kernels [46] were introduced in convolutional neural networks which significantly improve the accuracy of semantic segmentation. This study has employed the VGG-16, ResNet34, and InceptionV3 CNN models as an encoder for the semantic segmentation of cotton bolls and sky because of their state-of-art-performance in deep learning in agriculture [3,4,14,27,33,47,71]. For that, fully connected dense layers were removed from these CNN models and U-Net [51] like topology was formed using up-sampling layers so that the size of the output is the same as input which is the basic requirement of semantic segmentation task. A conceptual diagram of the CNN model used as an encoder for semantic segmentation is shown in Fig. 3. The transfer learning approach was used to fine-tune proposed CNN models as fine-tuning the CNN model with transfer learning is much faster than training the network from beginning with the randomly initialized weights. Transfer learning is a technique that is employed by transferring the weights of a pre-trained network and fine-tuning it for another task using new image datasets. This technique is advantageous when the image dataset for CNN training is small. The deep learning models used in this study are explained in the next subsections.

3.3.1. ResNet34

As the number of convolutional layers increases, backpropagation encounters the vanishing gradient problem, and the accuracy starts to saturate which degrades rapidly. He et al. [25] proposed a deep residual network also called ResNet which solves the vanishing gradient problem by introducing shortcut connections which are added to add the input x to the output after a few convolutional layers. In the present study, ResNet34 was used as an encoder which contains a 7 × 7 convolutional layer with 64 kernels, a 3 × 3 max pooling layer with stride 2 and 16 residual blocks.

3.3.2. VGG16

VGG16, a convolutional neural network proposed by Simonyan and Zisserman [57] is used as an encoder in this study which is made up of 13 convolutional layers. Each convolutional layer is followed by the batch normalization and the ReLU layers. To reduce the number of trainable parameters of the architecture, only 3 × 3 filter size was used in all convolutional layers. Max pooling layers were used between the convolutional layers to decrease the neural network dimension.

3.3.3. InceptionV3

InceptionV3 is a convolutional neural network architecture from the inception family [64–66] in which authors proposed the use of factorized convolutions, smaller convolutions, and asymmetric convolutions to reduce the number of trainable parameters and thus, the computational power. This version of Inception has five inception modules A, four inception modules B, and two inception modules C. All these inception modules are thoroughly explained in literature [66].

3.4. Training of convolutional neural networks models

For training convolutional neural networks, several hyperparameters need to be tuned to achieve better accuracy. Adam [34] is an iterative optimization algorithm that estimates the gradient mean and element-wise squared gradient of the gradient using the training dataset and then updates the internal parameters of a model using back-propagation. Instead of updating weights with the full amount, it is scaled by a factor called learning rate which is an important hyperparameter when training convolutional neural networks [39]. A small learning rate may result in poor generalization performance of the convolutional neural network [80] as a model may converge to some local minima and will also result

Fig. 2. Sample of augmented raw images and corresponding labeled images.

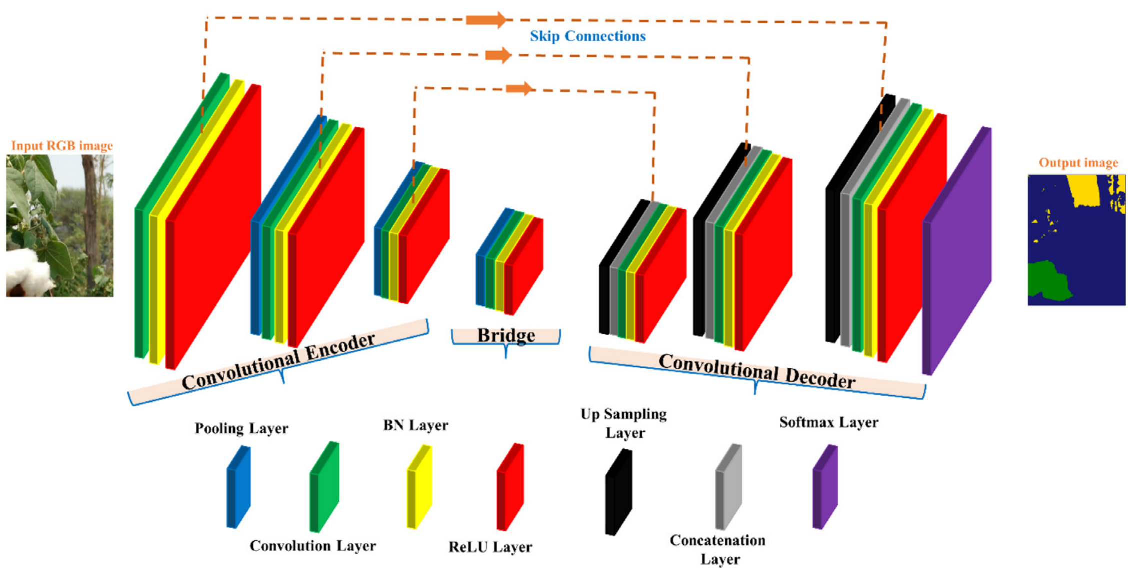

Fig. 3. Conceptual diagram of CNN model used as an encoder with skip connections.

in unnecessarily long training time as weights will be updated slowly. On the other hand, too large a learning rate may overshoot the desired minima such that they just osculate around the minima and will result in a sub-optimal set of weights. The number of epochs, another hyper-parameter of importance, is the number of times the learning algorithm will work through the entire training dataset. Another important parameter in case of convolutional neural network training is the batch size which is the number of images passes to the learning algorithm before internal model parameters are updated. Large batch size will lead to over-consumption of memory [79] whereas small batch size can have a significant regularization effect [78], and Kandel and Castelli [30] recommended small batch size with a low learning rate. For agricultural images, numerous researchers implemented Adam optimizer [10,22,74], a learning rate of 0.001 [22,35,84], trained model to one hundred epochs [2,35,68,86] and achieved impressive accuracies.

Therefore, in this study, an Adam optimizer with a learning rate of 0.001 was used for CNN models training. The models were trained for one hundred epochs with a small batch size of eight images for better generalization of models.

For semantic segmentation using CNN models, the presence of a strong class imbalance in the training dataset may result in sub-optimal performance by trained CNN [21]. As a large pixel ratio gap is obvious for cotton, sky, and background pixels in the image dataset of the present study, therefore to deal with the class imbalanced issue, the Dice coefficient as a loss function was used in this study. In the past, multiple researchers [7,45,67] successfully used the Dice coefficient as a loss function in semantic segmentation for agricultural images. The Dice coefficient loss function is defined by Eq. (1) [21].

$$\mathrm{Dice_{Loss}} = 1 - \frac{2 \times p_c^i \times q_c^i + \epsilon}{p_c^i + q_c^i + \epsilon} \tag{1}$$

Where, i = ith sample in training data; c = cth class; p_c^i = one hot encoder of ground truth; q_c^i = probability of class c for ith sample; $\epsilon = 1 \times 10^{-6}$

Table 1
Specifications of software and hardware.

Name	Specifications
GPU	1xTesla T4, having 2496 CUDA cores, compute 3.7, 16 GB (15.10 GB Usable) GDDR5 VRAM
CPU	1xsingle core hyperthreaded i.e. (1 core, 2 threads) Xeon (R) Processors @2.2Ghz, 56.3 MB Cache
RAM	25 GB Available
Disk	188 GB Available
Programming Language	Python
Development Environment	Jupyter notebook

Table 2
Evaluation metrics.

Metric	Formula
IoU	$\frac{TP_i}{TP_i + FN_i + FP_i}$
Precision	$\frac{TP_i}{TP_i + FP_i}$
Recall	$\frac{TP_i}{TP_i + FN_i}$
F1-score	$\frac{2 \times Precison_i \times Recall_i}{Precision_i \times Recall_i}$

Before the initialization of the training process, the image dataset was split into training, validation, and testing sets [35,61,86]. Out of 800 images, randomly 560 images (70%) for training, 160 images (20%) for validation, and 80 (10%) for testing purposes were used. Training of proposed models was implemented on the public Keras [16] library with TensorFlow [1] as the backend. Python programming language and Google Colaboratory [23] were used to train and validate the models. The overall specifications of Google Colaboratory are shown in Table 1.

3.5. Evaluation metrics for semantic segmentation

To evaluate the quality of segmentation results, the output segmented images from CNN models were compared with ground truth images and overlap between them was calculated. Intuitively, a good prediction maximizes the overlap between the predicted and ground truth. Numerous researchers [6,15,31,58,61,67] used F1-score, Intersection over Union (IoU), precision and recall metrics to evaluate the semantic segmentation accuracy in their study. In the present study, we used four well-known metrics: F1-score, IoU, precision, and recall to evaluate the semantic segmentation results quantitatively. Intersection-over-union (IoU) is an important accuracy measure that quantifies the percent overlap between predicted classes and ground truth classes and in the present study it was calculated separately for each class for better comparison of models' accuracy. For a particular class say *i*, precision measures the fraction of correctly classified pixels of class *i* among the total pixels classified as class *i* pixels. Similarly, recall measures the fraction of correctly classified pixels of class *i* among the actual number of pixels of class *i* present in the ground truth image. F1 score combines the precision and recall into a single parameter and a perfect model should have an F1 score of one [58]. A good F1 score means that model generates results with low false positives as well as a low false negative Table 2. shows the equations for these evaluation metrics. In these equations, true positive (TPi) are pixels belong to class *i* that is predicted correctly, false positive (FPi) are pixels that are predicted as class *i* pixels but does not belong to class i, false negative (FNi) are pixels that belong to class *i* but incorrectly classified as pixels of another class. To assess the segmentation speed, segmentation time (ST) was also measured.

4. Results and discussion

This section represents a quantitive evaluation of the trained CNN models for the detection of cotton bolls and sky.

4.1. Training results

Fig. 4 shows the F1-score, IoU score, and loss curves of the CNN models for each epoch of the training and validation process. After a certain number of epochs, for all models, the F1 score and IoU score was greater than 90% and loss value was less than 0.2, and beyond that, the training and validation learning curves show no significant improvement in terms of either F1-score, IoU score or loss, and therefore, training was stopped at one hundred epochs to prevent models from overfitting. Features map visualization helps to get intuitive about the features detected by CNN models explicitly which will help to understand the reasons for which the model might not be performing so well for some of the images and hence, considering that knowledge, fine-tuning of CNN model can be performed for better results Fig. 5. shows the features maps extracted at lower layers using trained CNN models for a better understanding of features learned by models. It can be observed that the different filters extracted different features using color, texture, and edges features of the input image.

4.2. Segmentation results

To evaluate the segmentation results, two test datasets were used in this study. The first dataset named the cotton-sky dataset was having 80 images of the cotton bolls having sky in the background. This dataset includes single as well as multiple instances of cotton bolls. Another dataset called the cotton boll dataset was having 73 images of the cotton bolls and the sky is not present in the background in this dataset Fig. 6. shows the semantic segmentation results of cotton bolls and sky. By visual observing, qualitatively, it can be observed that CNN models segment the cotton bolls and sky pixels successfully. But there are some misclassifications of pixels' classes as shown by the red square in Fig. 6. In the red square, cotton pixels are misclassified as sky pixels or background pixels, whereas the black square shows the cotton bolls which were partially detected. Some cotton bolls are completely missed by CNN models as shown by the red square in the fifth column of Fig. 6. To describe the performance of CNN models, a confusion matrix in which counts of predicted and actual results are presented in a tabular form is used in this study to represent the predictions of CNN models Fig. 7. shows the confusion matrices for trained CNN models of all three classes in which diagonal values represent the true positive Table 3.

As the dataset in the present study is moderately imbalanced for the cotton bolls class for obvious reasons, therefore, the precision-recall curve is plotted as shown in Fig. 8 for analyzing the models. As the area under the curves for each class for every model is higher which means models are predicting accurate results, as well as a majority of them, are positive results for each class.

The quantitative analysis of segmentation accuracy obtained by each model for the cotton -sky dataset in terms of IoU, F1-score, precision, and recall are presented in Table 3. The best value of each metric among all trained models is highlighted. For InceptionV3 model, IoU, F1-score and recall values for cotton bolls are 84.50%, 95.0% and 91.0%, respectively while for sky, these values are 80.67%, 96.0% and 91.0%, respectively. In terms of IoU for the sky, all three models perform equally well as the difference between IoU values is not so significant (< 1%). But, for the IoU value of cotton, the InceptionV3 model outperforms the

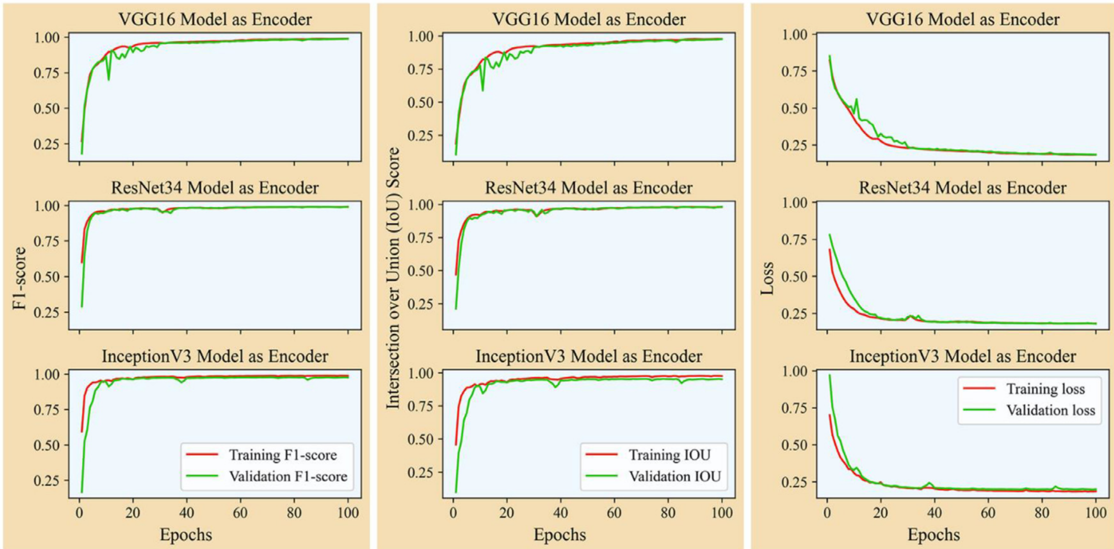

Fig. 4. Training and validation learning curves for the trained CNN models.

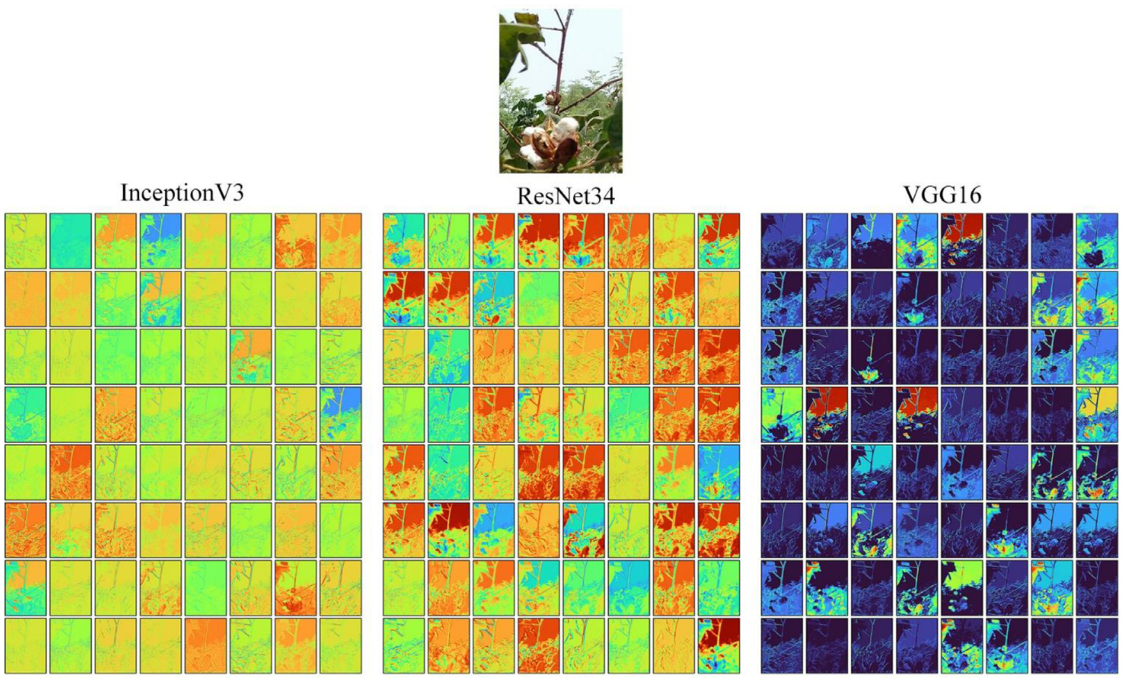

Fig. 5. Visualization of low-level feature maps using trained CNN models.

VGG16 model by 3.5%. Also, there is a significant difference (> 5%) for metrics F1-score and recall, in the case of the InceptionV3 model, as compared to VGG16 and ResNet34 models. In past, Li et al. [38] used convolutional networks for cotton bolls detection in which they consider the sky interference and achieved an IoU of 59.1% for sky interference images. In another study, Li et al. [37] proposed a method for in-field cotton detection via region-based semantic segmentation in which they achieved 73.5%, 77.3%, and 99.3% of IoU, specificity, and sensitivity values for sky interference images.

For the complementary purpose of this study, in addition to comparing trained models with image datasets consisting of the sky as well as cotton bolls images, trained models were also evaluated on the image dataset which was not having sky images but of cotton bolls only. Fig. 9 shows the semantic segmentation results of cotton bolls. By visual observing, qualitatively, it can be observed that trained CNN models segment the cotton bolls successfully. But few cotton pixels are misclassified as sky pixels which are highlighted using the red square in Fig. 9. Table 4 shows the quantitative results of segmentation accuracy obtained using trained models for segmentation of cotton bolls in terms of F1-score, precision, recall, and IoU scores, and the best values are highlighted.

It can be observed from Table 4 that all three trained models detect cotton bolls effectively as the IoU score for each model is above 90%. The highest IoU of 93.29% was achieved in the case of the InceptionV3 model with precision and recall values of 96.0% and 96.5%, respectively. Yeom et al. [83] proposed an open boll detection algorithm using ultra-fine spatial resolution UAV (Unmanned aerial vehicle) images and achieved maximum values of 92.5%, 96.6%, 95.6%, and 96.1% for

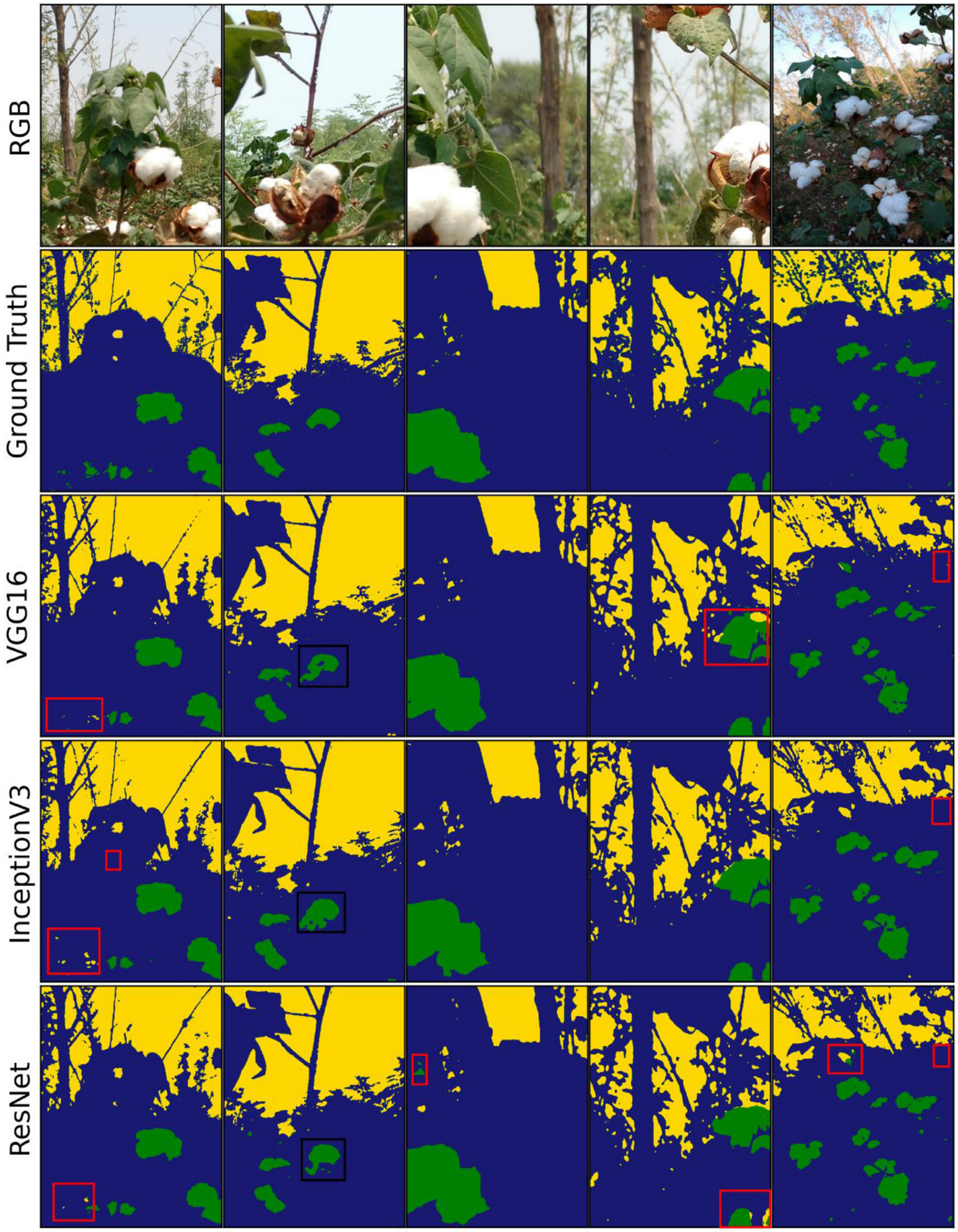

Fig. 6. Sematic segmentation results of cotton bolls and sky (In red square, cotton pixles are misclassified as sky pixels or background pixels, whereas the black square shows the cotton bolls which were detected partially.).

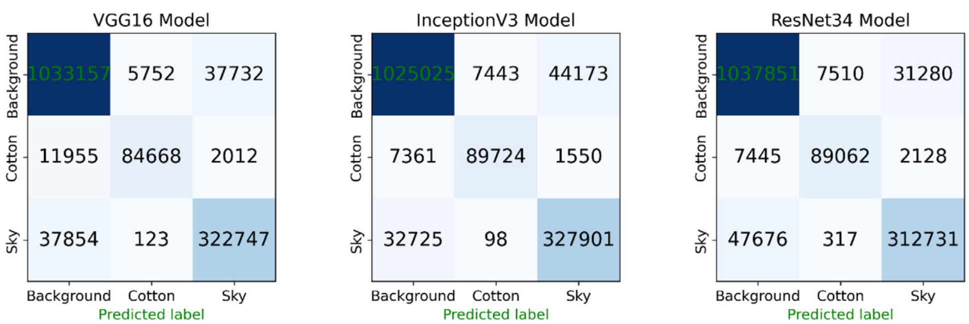

Fig. 7. Confusion matrix for the trained models.

Table 3
Evaluation of segmentation results obtained using models for cotton and sky images.

Evaluation Metrics	VGG16			InceptionV3			ResNet34		
	Background (%)	Cotton (%)	Sky (%)	Background (%)	Cotton (%)	Sky (%)	Background (%)	Cotton (%)	Sky (%)
IOU	91.72	81.01	80.59	91.78	**84.50**	**80.67**	91.70	83.65	79.34
F1-Score	96.0	90.0	89.0	96.0	**95.0**	**96.0**	96.0	91.0	88.0
Recall	96.0	86.0	89.0	95.0	**91.0**	**91.0**	96.0	90.0	87.0
Precision	95.0	**94.0**	89.0	96.0	92.0	88.0	95.0	92.0	**90.0**

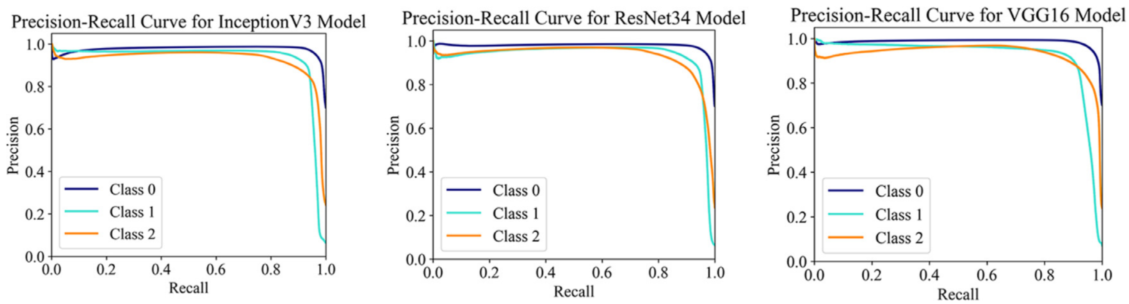

Fig. 8. Precision-recall curves for trained models. (*Class0: Background; Class1: Cotton; Class2: Sky*)

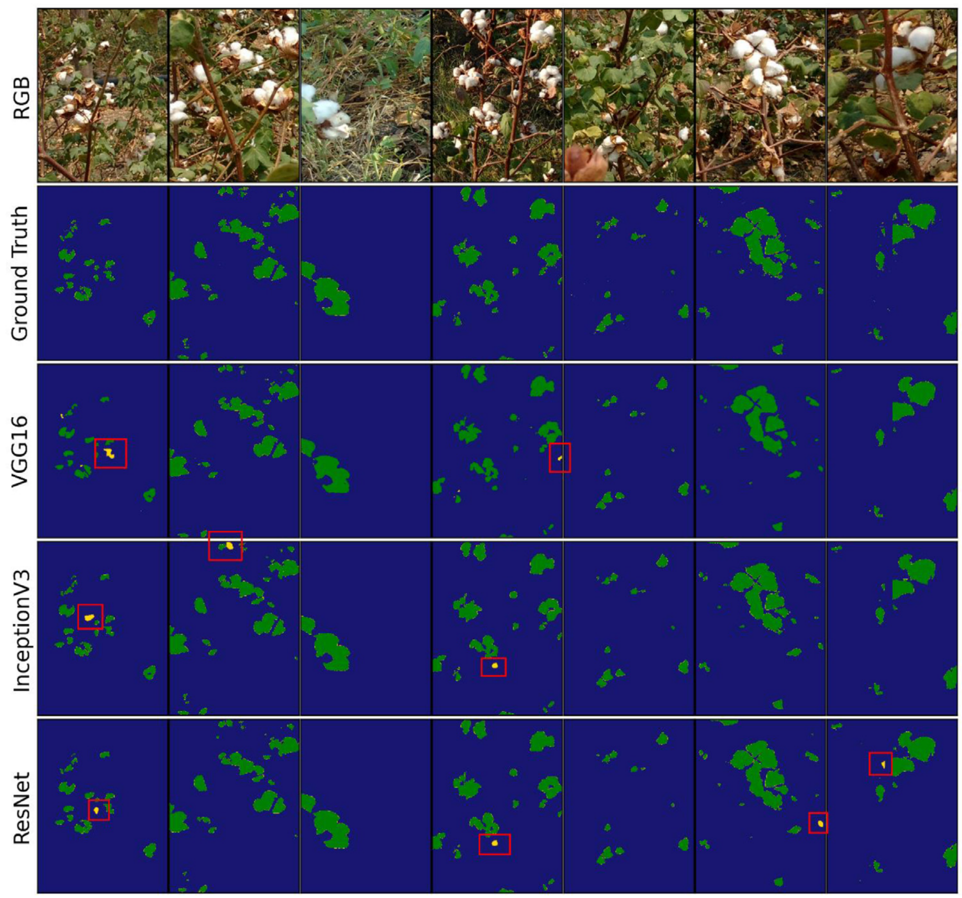

Fig. 9. Semantic segmentation results of cotton bolls (Red square highlight the cotton pixels which were classified as sky pixels).

Table 4
Evaluation of segmentation results obtained using models for cotton images only.

Evaluation Metrics	VGG16		InceptionV3		ResNet34	
	Background (%)	Cotton (%)	Background (%)	Cotton (%)	Background (%)	Cotton (%)
IOU	99.33	90.38	99.52	**93.29**	99.35	90.92
F1-Score	99.99	**98.0**	99.99	96.24	99.99	95.0
Recall	99.99	93.0	99.99	**96.5**	99.99	95.0
Precision	99.99	95.0	99.99	**96.0**	99.99	95.0

Table 5
Number of trainable parameters and processing time per image.

Metrics	VGG16	InceptionV3	ResNet34
Trainable Parameters	23.75 M	29.89 M	24.44 M
Segmentation time	1.34 s	1.07 s	1.02 s

IoU, precision, recall, and F1-measure respectively, whereas Singh et al. [58] used color features to segment cotton bolls and achieved precision and recall values of 97.53% and 88.69%, respectively.

Segmentation time (ST) is defined as the average time needed to segment a given number of images selected randomly from the dataset. For real-time applications of the CNN models, segmentation time is also an important parameter and should be considered along with other performance parameters. Therefore, in this study, segmentation time for semantic segmentation of cotton bolls and sky by each CNN model was calculated and shown in Table 5 along with the computational complexity of convolutional networks in terms of trainable parameters. It can be observed that there is no significant difference between segmentation time by InceptionV3 and ResNet34 models while InceptionV3 model shows higher segmentation performance.

4.3. Discussion of limitations

Trained CNN models in this paper can discriminate the pixels of cotton and sky successfully with an IoU score of greater than 80%. These models can be used in the vision system of cotton harvesting robots for the detection of open cotton bolls which are due for harvesting in a week. Despite these encouraging results, there is further space for improvements. Certain limitations of the models are discussed below.

1 In cotton fields, sometimes cotton bolls were clustered together and overlapped with each other [62]. The proposed models are unable to separate these overlapped cotton bolls.
2 Segmentation time considerably affects the time taken by a harvesting robot to complete one harvest cycle [73]. Though the segmentation time for the proposed models is less than 1.5 s, still there is a scope to reduce this time by developing models with lesser parameters.
3 For the real-time application of deep neural networks using resource-constrained mobile devices, for example, Raspberry Pi [49], a computationally efficient network with low memory is mandatory [22]. The size of trained neural networks in the present study is above 270 MB, which limits their use to personal computers only.

These limitations are part of our current ongoing research and are beyond the scope of the present paper.

5. Conclusions and future work

This study addressed one of the major challenges of in-field discrimination of cotton bolls from the sky using a fully convolutional neural network. To achieve this goal, three pre-trained convolutional neural networks namely VGG16, InceptionV3, and ResNet34 was used as an encoder. These CNN models were trained using the concept of transfer learning for which the cotton-sky image dataset was prepared and used as a testing and validation dataset. Trained CNN models were evaluated with different metrics to describe the effectiveness of the CNN models quantitively. For performance evaluation, these CNN models have been applied on the cotton-sky dataset as well as the cotton bolls dataset. For the cotton sky dataset, all three CNN models show IoU over 81% with precision and recall values of over 88%. For the cotton bolls dataset, all three models show IoU over 90% with precision and recall values above 95%. InceptionV3 model outperforms the other two models with IoU for cotton bolls and sky of above 84% and 80%, respectively and F1-score for cotton bolls and sky of above of 95% and 96%, respectively. The segmentation time of the InceptionV3 model was measured to be 1.07 s. Furthermore, for the cotton bolls dataset, InceptionV3 shows an IoU value of above 93% for cotton bolls. Hence, it can be concluded that the InceptionV3 model will segment cotton bolls and sky with greater accuracy and low error rates and can be used effectively for semantic segmentation of cotton bolls for harvesting it using robots.

In future extensions of this work, we will further investigate the use of deep learning models to improve the segmentation efficiency with the reduced number of trainable parameters and lower segmentation time.

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.

Acknowledgment

The financial support was received from the Indian Institute of Technology, Kharagpur, West Bengal, India.

References

[1] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D.G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, X. Zheng, TensorFlow: A system for large-scale machine learning, Business Opportunity 10 (2016) 265–283.
[2] A. Abdalla, H. Cen, L. Wan, R. Rashid, H. Weng, W. Zhou, Y. He, Fine-tuning convolutional neural network with transfer learning for semantic segmentation of ground-level oilseed rape images in a field with high weed pressure, Computers and Electronics in Agriculture 167 (2019) 105091, doi:10.1016/j.compag.2019.105091.
[3] H. Afzaal, A.A. Farooque, A.W. Schumann, N. Hussain, A. McKenzie-Gopsill, T. Esau, F. Abbas, B. Acharya, Detection of a potato disease (early blight) using artificial intelligence, Remote Sensing 13 (2021) 411–413 2021Page411, doi:10.3390/RS13030411.
[4] B.S. Anami, N.N. Malvade, S. Palaiah, Deep learning approach for recognition and classification of yield affecting paddy crop stresses using field images, Artificial Intelligence in Agriculture 4 (2020) 12–20, doi:10.1016/J.AIIA.2020.03.001.
[5] Apeer, 2021. Apeer [WWW Document]. URL https://www.apeer.com/home (accessed 9.14.21).
[6] A. Azizi, Y. Abbaspour-Gilandeh, E. Vannier, R. Dusséaux, T. Mseri-Gundoshmian, H.A. Moghaddam, Semantic segmentation: A modern approach for identifying soil clods in precision farming, Biosystems Engineering 196 (2020) 172–182, doi:10.1016/j.biosystemseng.2020.05.022.
[7] V. Badrinarayanan, A. Kendall, R. Cipolla, SegNet: A deep convolutional encoder-decoder architecture for image segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence 39 (2017) 2481–2495, doi:10.1109/TPAMI.2016.2644615.
[8] K. Bakhsh, N. Ahmad, M.A. Kamran, S. Hassan, Q. Abbas, R. Saeed, M.S. Hashmi, Occupational hazards and health cost of women cotton pickers in Pakistani Punjab, BMC Public Health 16 (2016) 961, doi:10.1186/s12889-016-3635-3.
[9] K. Bakhsh, N. Ahmad, S. Tabasum, S. Hassan, I. Hassan, Health hazards and adoption of personal protective equipment during cotton harvesting in Pakistan, Science of the Total Environment 598 (2017) 1058–1064, doi:10.1016/j.scitotenv.2017.04.043.
[10] W. Bao, X.X. Yang, D. Liang, G. Hu, X.X. Yang, Lightweight convolutional neural network model for field wheat ear disease identification, Computers and Electronics in Agriculture 189 (2021) 106367, doi:10.1016/j.compag.2021.106367.
[11] R. Barth, J. IJsselmuiden, J. Hemming, E.J. Van Henten, Synthetic bootstrapping of convolutional neural networks for semantic plant part segmentation, Computers and Electronics in Agriculture 161 (2019) 291–304, doi:10.1016/j.compag.2017.11.040.
[12] J. Brown, S. Sukkarieh, Design and evaluation of a modular robotic plum harvesting system utilizing soft components, Journal of Field Robotics 38 (2021) 289–306, doi:10.1002/rob.21987.
[13] D.M. Bulanon, T. Kataoka, Y. Ota, T. Hiroma, AE-Automation and emerging technologies: A segmentation algorithm for the automatic recognition of fuji apples at harvest, Biosystems Engineering 83 (2002) 405–412, doi:10.1006/bioe.2002.0132.
[14] C. Chen, B. Li, J. Liu, T. Bao, N. Ren, Monocular positioning of sweet peppers: An instance segmentation approach for harvest robots, Biosystems Engineering 196 (2020) 15–28, doi:10.1016/j.biosystemseng.2020.05.005.
[15] Z. Chen, D. Ting, R. Newbury, C. Chen, Semantic segmentation for partially occluded apple trees based on deep learning, Computers and Electronics in Agriculture 181 (2021) 105952, doi:10.1016/j.compag.2020.105952.
[16] Chollet, F., 2015. Keras [WWW Document]. URL https://github.com/keras-team/keras (accessed 9.4.21).
[17] K.H. Dammer, Real-time variable-rate herbicide application for weed control in carrots, Weed Research 56 (2016) 237–246, doi:10.1111/wre.12205.
[18] A. De Preter, J. Anthonis, J. De Baerdemaeker, Development of a robot for harvesting strawberries, IFAC-PapersOnLine 51 (2018) 14–19, doi:10.1016/j.ifacol.2018.08.054.

[19] A. Feng, J. Zhou, E.D. Vories, K.A. Sudduth, M. Zhang, Yield estimation in cotton using UAV-based multi-sensor imagery, Biosystems Engineering 193 (2020) 101–114, doi:10.1016/j.biosystemseng.2020.02.014.

[20] K.G. Fue, W.M. Porter, E.M. Barnes, G.C. Rains, An extensive review of mobile agricultural robotics for field operations: Focus on cotton harvesting, AgriEngineering. (2020), doi:10.3390/agriengineering2010010.

[21] H. Fujii, H. Tanaka, M. Ikeuchi, K. Hotta, X-net with different loss functions for cell image segmentation, in: Proceedings of the 2021 IEEE/CVF conference on computer vision and pattern recognition workshops (CVPRW), 2021, pp. 3788–3795, doi:10.1109/CVPRW53098.2021.00420.

[22] V. Gonzalez-Huitron, J.A. León-Borges, A.E. Rodriguez-Mata, L.E. Amabilis-Sosa, B. Ramírez-Pereda, H. Rodriguez, Disease detection in tomato leaves via CNN with lightweight architectures implemented in Raspberry Pi 4, Computers and Electronics in Agriculture 181 (2021) 105951, doi:10.1016/J.COMPAG.2020.105951.

[23] Google Colaboratory, 2021. Frequebtly Asked Questions [WWW Document]. Google. URL https://research.google.com/colaboratory/faq.html (accessed 9.5.21).

[24] E. Hamuda, B. Mc Ginley, M. Glavin, E. Jones, Improved image processing-based crop detection using Kalman filtering and the Hungarian algorithm, Computers and Electronics in Agriculture 148 (2018) 37–44, doi:10.1016/j.compag.2018.02.027.

[25] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE computer society conference on computer vision and pattern recognition, 2016 2016-december, 770–778, doi:10.1109/CVPR.2016.90.

[26] M.H. Hu, Q.L. Dong, B.L. Liu, L.Q. Pan, J. Walshaw, Image segmentation of bananas in a crate using a multiple threshold method, Journal of Food Process Engineering 39 (2016) 427–432, doi:10.1111/jfpe.12233.

[27] N. Hussain, A.A. Farooque, A.W. Schumann, F. Abbas, B. Acharya, A. McKenzie-Gopsill, R. Barrett, H. Afzaal, Q.U. Zaman, M.J.M. Cheema, Application of deep learning to detect Lamb's quarters (Chenopodium album L.) in potato fields of Atlantic Canada, Computers and Electronics in Agriculture 182 (2021) 106040, doi:10.1016/J.COMPAG.2021.106040.

[28] W. Ji, D. Zhao, F. Cheng, B. Xu, Y. Zhang, J. Wang, Automatic recognition vision system guided for apple harvesting robot, Computers & Electrical Engineering 38 (2012) 1186–1195, doi:10.1016/J.COMPELECENG.2011.11.005.

[29] L. Jin-Shuai, L. Hui-Cheng, J. Zhen-Hong, Image segmentation of cotton based on YCbCr color space and Fisher discrimination analysis, Acta Agronomica Sinica 37 (2011) 1274–1279.

[30] I. Kandel, M. Castelli, The effect of batch size on the generalizability of the convolutional neural networks on a histopathology dataset, ICT Express 6 (2020) 312–315, doi:10.1016/J.ICTE.2020.04.010.

[31] R. Kestur, A. Meduri, O. Narasipura, MangoNet: A deep semantic segmentation architecture for a method to detect and count mangoes in an open orchard, Engineering Applications of Artificial Intelligence 77 (2019) 59–69, doi:10.1016/j.engappai.2018.09.011.

[32] M. Khan, H.Z. Mahmood, C.A. Damalas, Pesticide use and risk perceptions among farmers in the cotton belt of Punjab, Pakistan, Crop Protection 67 (2015) 184–190, doi:10.1016/j.cropro.2014.10.013.

[33] M. Khanramaki, E. Askari Asli-Ardeh, E. Kozegar, Citrus pests classification using an ensemble of deep learning models, Computers and Electronics in Agriculture 186 (2021) 106192, doi:10.1016/j.compag.2021.106192.

[34] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, in: Proceedings of the 3rd international conference on learning representations, ICLR 2015 - conference track proceedings, 2014.

[35] S. Kolhar, J. Jagtap, Convolutional neural network based encoder-decoder architectures for semantic segmentation of plants, Ecological Informatics 64 (2021) 101373, doi:10.1016/j.ecoinf.2021.101373.

[36] N. Kondo, M. Monta, T. Fujiura, Fruit harvesting robots in Japan, Advances in Space Research 18 (1996) 181–184, doi:10.1016/0273-1177(95)00806-P.

[37] Y. Li, Z. Cao, H. Lu, Y. Xiao, Y. Zhu, A.B. Cremers, In-field cotton detection via region-based semantic image segmentation, Computers and Electronics in Agriculture 127 (2016) 475–486, doi:10.1016/j.compag.2016.07.006.

[38] Y. Li, Z. Cao, Y. Xiao, A.B. Cremers, DeepCotton: in-field cotton segmentation using deep fully convolutional network, Journal of Electronic Imaging 26 (2017) 1–14, doi:10.1117/1.JEI.26.5.053028.

[39] J. Liang, Y. Xu, C. Bao, Y. Quan, H. Ji, Barzilai–Borwein-based adaptive learning rate for deep learning, Pattern Recognition Letters 128 (2019) 197–203, doi:10.1016/J.PATREC.2019.08.029.

[40] L. Luo, Y. Tang, X. Zou, M. Ye, W. Feng, G. Li, Vision-based extraction of spatial information in grape clusters for harvesting robots, Biosystems Engineering 151 (2016) 90–104, doi:10.1016/j.biosystemseng.2016.08.026.

[41] Y. Majeed, J. Zhang, X. Zhang, L. Fu, M. Karkee, Q. Zhang, M.D. Whiting, Apple tree trunk and branch segmentation for automatic trellis training using convolutional neural network based semantic segmentation, IFAC-PapersOnLine 51 (2018) 75–80, doi:10.1016/j.ifacol.2018.08.064.

[42] M. Makky, P. Soni, Development of an automatic grading machine for oil palm fresh fruits bunches (FFBs) based on machine vision, Computers and Electronics in Agriculture 93 (2013) 129–139, doi:10.1016/J.COMPAG.2013.02.008.

[43] C.R. Mehta, N.S. Chandel, P.C. Jena, J. Anamika, Indian agriculture counting on farm mechanization. AMA, agricultural mechanization in Asia, Africa and Latin America 50 (2019) 84–89.

[44] W. Mulan, J. Wei, J. Yuan, K. Xu, A research for intelligent cotton picking robot based on machine vision, in: Proceedings of the 2008 international conference on information and automation. presented at the 2008 international conference on information and automation (ICIA), Changsha, China, IEEE, 2008, pp. 800–803, doi:10.1109/ICINFA.2008.4608107.

[45] L.C. Ngugi, M. Abelwahab, M. Abo-Zahhad, Tomato leaf segmentation algorithms for mobile phone applications using deep learning, Computers and Electronics in Agriculture 178 (2020) 105788, doi:10.1016/J.COMPAG.2020.105788.

[46] C. Peng, X. Zhang, G. Yu, G. Luo, J. Sun, Large kernel matters - Improve semantic segmentation by global convolutional network, in: Proceedings of the - 30th IEEE conference on computer vision and pattern recognition, CVPR 2017, 2017 2017-January, 1743–1751, doi:10.1109/CVPR.2017.189.

[47] A. Picon, A. Alvarez-Gila, M. Seitz, A. Ortiz-Barredo, J. Echazarra, A. Johannes, Deep convolutional neural networks for mobile capture device-based crop disease classification in the wild, Computers and Electronics in Agriculture 161 (2019) 280–290, doi:10.1016/j.compag.2018.04.002.

[48] P. Pinheiro, R. Collobert, Recurrent convolutional neural networks for scene labeling, in: Proceedings of the 31st international conference on machine learning. PMLR, 2014, pp. 82–90.

[49] Raspberry Pi, 2022. Teach, learn, and make with raspberry Pi [WWW Document]. Raspberry Pi. URL https://www.raspberrypi.org/ (accessed 2.20.22).

[50] J.P. Rodríguez, D.C. Corrales, J.N. Aubertot, J.C. Corrales, A computer vision system for automatic cherry beans detection on coffee trees, Pattern Recognition Letters 136 (2020) 142–153, doi:10.1016/j.patrec.2020.05.034.

[51] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional networks for biomedical image segmentation, Lecture Notes in Computer Science 9351 (2015) 234–241 (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), doi:10.1007/978-3-319-24574-4_28.

[52] A. Roshanianfard, N. Noguchi, Pumpkin harvesting robotic end-effector, Computers and Electronics in Agriculture 174 (2020) 105503, doi:10.1016/j.compag.2020.105503.

[53] S. Sabzi, Y. Abbaspour-Gilandeh, H. Javadikia, Machine vision system for the automatic segmentation of plants under different lighting conditions, Biosystems Engineering 161 (2017) 157–173, doi:10.1016/j.biosystemseng.2017.06.021.

[54] G. Samseemoung, P. Soni, P. Suwan, Development of a variable rate chemical sprayer for monitoring diseases and pests infestation in coconut plantations, Agriculture (2017), doi:10.3390/agriculture7100089.

[55] S.M. Seidu, Growth and instability in cotton cultivation in Northern India, EA 63 (2018), doi:10.30954/0424-2513.2.2018.20.

[56] A. Silwal, J.R. Davidson, M. Karkee, C. Mo, Q. Zhang, K. Lewis, Design, integration, and field evaluation of a robotic apple harvester, Journal of Field Robotics 34 (2017) 1140–1159, doi:10.1002/rob.21715.

[57] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, in: Proceedings of the 3rd international conference on learning representations, ICLR 2015 - conference track proceedings, 2014.

[58] N. Singh, V.K. Tewari, P.K. Biswas, C.M. Pareek, L.K. Dhruw, Image processing algorithms for in-field cotton boll detection in natural lighting conditions, Artificial Intelligence in Agriculture 5 (2021) 142–156, doi:10.1016/j.aiia.2021.07.002.

[59] V. Singh, Sunflower leaf diseases detection using image segmentation based on particle swarm optimization, Artificial Intelligence in Agriculture 3 (2019) 62–68, doi:10.1016/j.aiia.2019.09.002.

[60] C.E. Snipes, C.C. Baskin, Influence of early defoliation on cotton yield, seed quality, and fiber properties, Field Crops Research 37 (1994) 137–143, doi:10.1016/0378-4290(94)90042-6.

[61] K. Sun, X. Wang, S. Liu, C.H. Liu, Apple, peach, and pear flower detection using semantic segmentation network and shape constraint level set, Computers and Electronics in Agriculture 185 (2021) 106150, doi:10.1016/j.compag.2021.106150.

[62] S. Sun, C. Li, A.H. Paterson, P.W. Chee, J.S. Robertson, Image processing algorithms for infield single cotton boll counting and yield prediction, Computers and Electronics in Agriculture 166 (2019) 104976, doi:10.1016/j.compag.2019.104976.

[63] S. Sunoj, S.N. Subhashree, S. Dharani, C. Igathinathane, J.G. Franco, R.E. Mallinger, J.R. Prasifka, D. Archer, Sunflower floral dimension measurements using digital image processing, Computers and Electronics in Agriculture 151 (2018) 403–415, doi:10.1016/j.compag.2018.06.026.

[64] C. Szegedy, S. Ioffe, V. Vanhoucke, A. Alemi, Inception-v4, inception-ResNet and the impact of residual connections on learning, in: Proceedings of the 31st AAAI conference on artificial intelligence, 2017, AAAI, 2016, pp. 4278–4284.

[65] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: Proceedings of the IEEE computer society conference on computer vision and pattern recognition 07-12-June, 2014, pp. 1–9.

[66] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: Proceedings of the IEEE computer society conference on computer vision and pattern recognition 2016-Decem, 2015, pp. 2818–2826.

[67] H. Tang, B. Wang, X. Chen, Deep learning techniques for automatic butterfly segmentation in ecological images, Computers and Electronics in Agriculture 178 (2020) 105739, doi:10.1016/j.compag.2020.105739.

[68] L.M. Tassis, J.E. Tozzi de Souza, R.A. Krohling, A deep learning approach combining instance and semantic segmentation to identify diseases and pests of coffee leaves from in-field images, Computers and Electronics in Agriculture 186 (2021) 106191, doi:10.1016/j.compag.2021.106191.

[69] D. Tedesco-Oliveira, R. Pereira da Silva, W. Maldonado, C. Zerbato, Convolutional neural networks in predicting cotton yield from images of commercial fields, Computers and Electronics in Agriculture 171 (2020) 105307, doi:10.1016/j.compag.2020.105307.

[70] V.K. Tewari, C.M. Pareek, G. Lal, L.K. Dhruw, N. Singh, Image processing based real-time variable-rate chemical spraying system for disease control in paddy crop, Artificial Intelligence in Agriculture 4 (2020) 21–30, doi:10.1016/j.aiia.2020.01.002.

[71] K. Thenmozhi, U. Srinivasulu Reddy, Crop pest classification based on deep convolutional neural network and transfer learning, Computers and Electronics in Agriculture 164 (2019) 104906, doi:10.1016/j.compag.2019.104906.
[72] E.J. van Henten, J. Hemming, B.A.J. van Tuijl, J.G. Kornet, J. Meuleman, J. Bontsema, E.A. van Os, An autonomous robot for harvesting cucumbers in greenhouses, Autonomous Robots 13 (2002) 241–258, doi:10.1023/A:1020568125418.
[73] E.J. Van Henten, B.A.J. Van Tuijl, J. Hemming, J.G. Kornet, J. Bontsema, E.A. Van Os, Field test of an autonomous cucumber picking robot, Biosystems Engineering 86 (2003) 305–313, doi:10.1016/j.biosystemseng.2003.08.002.
[74] A. Waheed, M. Goyal, D. Gupta, A. Khanna, A.E. Hassanien, H.M. Pandey, An optimized dense convolutional neural network model for disease recognition and classification in corn leaf, Computers and Electronics in Agriculture 175 (2020) 105456, doi:10.1016/J.COMPAG.2020.105456.
[75] F. Wang, J. Zheng, X. Tian, J. Wang, L. Niu, W. Feng, An automatic sorting system for fresh white button mushrooms based on image processing, Computers and Electronics in Agriculture 151 (2018) 416–425, doi:10.1016/j.compag.2018.06.022.
[76] S. Wang, Y. Li, J. Yuan, L. Song, Xinghua Liu, Xuemei Liu, Recognition of cotton growth period for precise spraying based on convolution neural network, Information Processing in Agriculture 8 (2021) 219–231, doi:10.1016/j.inpa.2020.05.001.
[77] Y. Wang, X. Zhu, C. Ji, Machine vision based cotton recognition for cotton harvesting robot. Computer and computing technologies in agriculture, volume II. CCTA 2007, The International Federation for Information Processing 259 (2008) 1421–1425.
[78] D.R. Wilson, T.R. Martinez, The general inefficiency of batch training for gradient descent learning, Neural Networks 16 (2003) 1429–1451, doi:10.1016/S0893-6080(03)00138-2.
[79] W. Xie, S. Wei, Z. Zheng, D. Yang, A CNN-based lightweight ensemble model for detecting defective carrots, Biosystems Engineering 208 (2021) 287–299, doi:10.1016/j.biosystemseng.2021.06.008.
[80] Xing, C., Arpit, D., Tsirigotis, C., Bengio, Y., 2018. A walk with SGD.
[81] Y. Xiong, Y. Ge, P.J. From, An obstacle separation method for robotic picking of fruits in clusters, Computers and Electronics in Agriculture 175 (2020) 105397, doi:10.1016/j.compag.2020.105397.
[82] L. Xu, Y. Li, J. Xu, L. Guo, Two-level attention and score consistency network for plant segmentation, Computers and Electronics in Agriculture 170 (2020) 105281, doi:10.1016/j.compag.2020.105281.
[83] J. Yeom, J. Jung, A. Chang, M. Maeda, J. Landivar, Automated open cotton boll detection for yield estimation using unmanned aircraft vehicle (UAV) data, Remote Sensing (2018), doi:10.3390/rs10121895.
[84] L. Zabawa, A. Kicherer, L. Klingbeil, R. Töpfer, H. Kuhlmann, R. Roscher, Counting of grapevine berries in images via semantic segmentation using convolutional neural networks, ISPRS Journal of Photogrammetry and Remote Sensing 164 (2020) 73–83, doi:10.1016/j.isprsjprs.2020.04.002.
[85] K. Zhang, K. Lammers, P. Chu, Z. Li, R. Lu, System design and control of an apple harvesting robot, Mechatronics 79 (2021) 102644, doi:10.1016/j.mechatronics.2021.102644.
[86] K. Zou, X. Chen, Y. Wang, C. Zhang, F. Zhang, A modified U-Net with a specific data argumentation method for semantic segmentation of weed images in the field, Computers and Electronics in Agriculture 187 (2021) 106242, doi:10.1016/j.compag.2021.106242.