Technological University Dublin
ARROW@TU Dublin

Articles

School of Computer Sciences

2023

Towards Automated Weed Detection Through Two-Stage Semantic Segmentation of Tobacco and Weed Pixels in Aerial Imagery

S. Imran Moazzam

Umar S. Khan

Waqar Qureshi

See next page for additional authors

Follow this and additional works at: https://arrow.tudublin.ie/scschcomart

Part of the Computer Engineering Commons

This work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.
Funder: Higher Education Commission of Pakistan and the National center for Robotics and Automation (DF-1009–31). We thank Pakistan Tobacco Company for helping us find farms for data collection.

Authors
S. Imran Moazzam, Umar S. Khan, Waqar Qureshi, Tahir Nawaz, and Faraz Kunwar

Smart Agricultural Technology 4 (2023) 100142

Contents lists available at ScienceDirect

Smart Agricultural Technology

journal homepage: www.journals.elsevier.com/smart-agricultural-technology

Towards automated weed detection through two-stage semantic segmentation of tobacco and weed pixels in aerial Imagery

S. Imran Moazzam [a,b], Umar S. Khan [a,b], Waqar S. Qureshi [a,c,*], Tahir Nawaz [a,b], Faraz Kunwar [a]

[a] National University of Sciences and Technology H-12, Islamabad, Pakistan
[b] National Centre of Robotics and Automation (NCRA), Robot Design and Development Lab, National University of Sciences and Technology H-12, Islamabad, Pakistan
[c] Department of Computer Science, Technological University Dublin, Dublin, Ireland

ARTICLE INFO

Editor: Stephen Symons

Keywords:
Semantic segmentation
Tobacco weed detection
Tobacco dataset
Tobacco weed classification
Efficient semantic segmentation

ABSTRACT

In precision farming, weed detection is required for precise weedicide application, and the detection of tobacco crops is necessary for pesticide application on tobacco leaves. Automated accurate detection of tobacco and weeds through aerial visual cues holds promise. Precise weed detection in crop field imagery can be treated as a semantic segmentation problem. Many image processing, classical machine learning, and deep learning-based approaches have been devised in the past, out of which deep learning-based techniques promise better accuracies for semantic segmentation, i.e., pixel-level classification. We present a new method that improves the precision of pixel-level inter-class classification of the crop and the weed pixels. The technique applies semantic segmentation in two stages. In stage I, a binary pixel-level classifier is developed to segment background and vegetation. In stage II, a three-class pixel-level classifier is designed to classify background, weeds, and tobacco. The output of the first stage is the input of the second stage. To test our designed classifier, a new tobacco crop aerial dataset was captured and manually labeled pixel-wise. The two-stage semantic segmentation architecture has shown better tobacco and weeds pixel-level classification precision. The intersection over union (IOU) for the tobacco crop was improved from 0.67 to 0.85, and IOU for weeds enhanced from 0.76 to 0.91 with the new approach compared to the traditional one-stage semantic segmentation application. We observe that in stage I shallower, a smaller semantic segmentation model is enough compared to stage II, where a segmentation network with more neurons serves the purpose of good detection.

1. Introduction

Tobacco (Nicotiana tabacum) is a cash crop that highly depends on agrochemicals. High usage of pesticides and insecticides causes soil pollution. At the same time, for the safety of this crop, the tobacco industry considers the use of agrochemicals highly essential [17]. In Pakistan, tobacco farmers fail to avoid agrochemical exposure due to conventional pest/weed control methods. Human health and environmental concerns exist due to overapplication or uniform broadcast spraying on the whole field without distinguishing crop, weed, and background. Intelligent application of agrochemicals is required by detecting crop plants, weeds, and soil background.

Tobacco yield could be affected by insects and weeds present in the field that could share essential nutrients needed by tobacco. Weed is one of the major issues being faced by agriculture as it can spread quickly and decrease crop yield and quality. Weed could compete with crops by sharing water, sunlight, growing space, and nutrition [6]. Weedicide application is mainly done to counter weeds on farms, and it requires time, has a financial burden, and may impact people's health, soil, and environment [5]. Due to an increase in labor costs and health and environmental concerns, automated weed control is desirable. An automated selective spraying approach could minimize the use of herbicides [12]. Efficient detection of crop plants is essential for selective pesticide spray application on crops. The first important step to developing an autonomous weed detection system that correctly recognizes weeds.

In the past various image processing and machine learning, including deep learning-based approaches, have been applied, out of which deep learning-based techniques have shown better performance on vision problems. Their performance improves further with the availability of

This work is funded by the Higher Education Commission of Pakistan and The National center of Robotics and Automation (DF-1009–31).
* Corresponding author.
E-mail address: waqar.qureshi@tudublin.ie (W.S. Qureshi).

https://doi.org/10.1016/j.atech.2022.100142
Received 5 May 2022; Received in revised form 20 November 2022; Accepted 23 November 2022
Available online 1 December 2022
2772-3755/© 2022 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

more data in vision tasks. Crop and weeds have many similar characteristics, and extracting discriminative features is challenging using the classical machine learning approach, which can be addressed by applying deep learning due to its enhanced feature learning capabilities. Next, we briefly discuss limitations found in the literature for crop and weed classification.

1.1. Image processing and machine learning-based methods for background removal

To differentiate between the background and green vegetables, Milioto et al. [14] applied thresholding using the Otsu-adaptive method on Normalized Difference Vegetation Index (NDVI) to remove the background. Espejo-Garcia et al. [4] and Andrea et al. [3] applied normalization on RGB channels before applying Otsu's thresholding. Knoll et al. [11] used HSV color space instead of RGB. Le et al. [13] used Excess Green minus Excess Red Indices (ExG-ExR) method for background removal. Jiang et al. [8] applied histogram equalization to remove the background and enhance image contrast. However, histogram equalization also improves the contrast of shadows; therefore, vegetation in shadows is less likely to be detected. [2] applied several morphological operations and thresholding methods to separate the soil and vegetation. Image processing-based methods that use different color spaces and pre and post-processing before applying a threshold have been used extensively; however, choosing an optimum threshold is challenging for a real-world aerial image with variation in vegetation color due to non-linear variation in lighting conditions.

Sabzi [21] presented a machine-learning approach for background removal. He used eight shape features, eight texture features, five moment-invariant features, and thirteen color features. They extensively analyzed ant colony, simulated annealing, genetic algorithm, cultural algorithm, linear discriminant analysis (LDA), support vector machines (SVM), and random forest classifiers to distinguish between crops and weeds. In a classical machine learning approach, researchers have experimented with different hand-crafted features with different classifiers algorithms for different crop types. For corn crops, weeds are detected from hyperspectral images by applying SVM [9]. Wendel & Underwood [22] used SVM and LDA for plant classification. Different shape features are extracted to distinguish between narrow and broad-leafed weeds, and the feature vectors are evaluated using a single-layer perceptron [7]. Handcrafting of features for artificial intelligence (AI) system that detects crop and weed is challenging and give rise to self-learning feature detection techniques through deep learning approaches.

1.2. Crop and weed classification using an object-detection-based approach

In the object-detection-based crop and weed classification approach, a vegetation blob or a bounding box within an image is classified as either a crop or a weed. In this case, labeling individual plants as a weed or crop plants is important. Many researchers have used this approach using image-level annotations and bounding box annotations, such as Nkemelu et al. [15] and Partel, et al. [16]. The plant-based classification is faster than semantic segmentation; however, it has a localization problem when the weeds are close or occluded with the crop plant. The weeds near the plant's roots (occluding the plant) are more hazardous to crop health. So, to get an accurate detection of every pixel in an image, semantic segmentation is a better option.

1.3. Crop and weed classification using a semantic segmentation approach

Another approach in convolution neural networks is weed mapping, based on semantic segmentation. Abdalla et al., [1] and I. Sa et al. [20] approach to crop detection maps every pixel of weed and crop in the field. Semantic segmentation approaches are computationally expensive as appose to object-detection-based approaches. Another limitation of the semantic segmentation-based weed and crop detection approach is poor inter-class pixel classification due to similar texture and color features. Previous application of semantic segmentation has shown effective removal of the background while the crop and the weed inter-class classification accuracy is still a challenge. We have addressed this limitation in this research by choosing a 2-stage semantic segmentation application.

Table 1
Tobacco Dataset Drone fly campaigns.

Field No.	Attribute / No. of images of 480 × 352 resolution	Geographical Coordinates Latitude/ Longitude	Timing Around/ Date Captured	Soil Condition/ Sunlight Condition
147	Train / 864	34;19;54.998/ 71;57;21.032	2:30pm/09 April 2021	Wet/ Sunny
119	Test / 936	34;18;22.110/ 72;03;6.744	3:30pm/07 April 2021	Dry and Wet/ Sunny
120	Test / 120	34;18;22.052/ 72;03;6.744	3:38pm/07 April 2021	Wet/ Sunny
133	Test / 120	34;19;58.741/ 71;57;9.261	5:52pm/07 April 2021	Semi Wet/ Sunny
134	Test / 120	34;19;55.400/ 71;57;19.676	6:33pm/07 April 2021	Wet/ Sunny
154	Test / 120	34;19;49.984/ 71;57;15.300	3:27pm/09 April 2021	Semi Wet/ Cloudy
163	Test / 120	34;19;57.650/ 71;57;17.873	3:43pm/09 April 2021	Dry and Wet/ Sunny
171	Test / 120	34;19;58.666/ 71;57;18.553	3:59pm/09 April 2021	Dry and Wet / Sunny

1.4. Contributions

This research proposes a new method of automated weed detection using a two-stage semantic segmentation method to increase the accuracy of pixel-level inter-class classification of tobacco and weed pixels. The algorithm is tested on an aerial early-stage tobacco crop dataset captured during the 2021 crop season.

2. Material and methods

This section provides datasets details, the proposed method, and its implementation details.

2.1. Tobacco dataset

We have acquired a new tobacco-weed dataset using a Mavic Mini drone. Eight fields of tobacco crops are captured in Mardan, Khyber Pakhtunkhwa, Pakistan. At different growth stages, these eight fields are captured at a crop age of 15 to 40 days approximately. Images are captured at 1920 × 1080-pixel resolution; due to system memory limitations, we have cropped non-overlapping images of resolution 480 × 352 for processing. This image patch cropping is implemented using a code that reads images and creates non-overlapping tile images using two nested loops; the respective annotation images are also cropped simultaneously.

Images from an entire field are used in training, and images from the rest of the seven areas are used in testing to see the fitness of our algorithm. Table 1 lists these eight drone fly campaigns. Dataset is captured at an average altitude of 4 m with a ground sampling distance of 0.1 cm/pixel. Fig. 1 shows an example image of a tobacco dataset in 1920 × 1080-pixel resolution. Images are labeled manually; background, crop, and weed have label values of 0, 1, and 2, respectively. We have publicly made this tobacco-weed dataset available on this link (S. [18]).

Fig. 1. Tobacco dataset images from different fields infested with weeds, different lighting conditions, and plant growth stages are obvious.

2.2. Weed mapping using 2-stage semantic segmentation

We attempt crop-weed-background detection in two stages. Original 1080P images are converted into patches of size 480 × 352 for training and testing. The background is distinctive to the other two classes (Crop and Weed) and so to maximize efficiency, in stage-I, we have applied semantic segmentation to distinguish between background and vegetation. Detected background pixels in stage-I are replaced by zeroes to simplify data. Then this simplified data is passed through stage-II for the final classification of crop, weed, and background. Semantic segmentation has the greatest computational complexity of all deep learning techniques and its application in two stages has increased the segmentation accuracy as well as computational complexity. To reduce the complexity of this proposed approach we have done a selection of simpler encoder-decoder segmentation networks in the first stage. As we know that the background-vegetation detection problem is simpler as compared to the crop-weed detection problem, we have used a simpler semantic segmentation model in stage-I as compared to stage II to reduce complexity while achieving better accuracy.

Flow diagram of our proposed two-stage semantic segmentation application and traditional way of semantic segmentation application are shown in Fig. 2. Original image to patch conversion is shown in Fig. 3. Patches of size 480 × 352 are used in training and testing. Since the last 24 rows in original 1080P images are covered in the next frame, as each image cover 80% of the previous frame so it does not affect the overall coverage of the field. In addition to a two-time semantic segmentation application, our method has another core requirement, which is the choice of a checkpoint where vegetation detection of the model is maximum in stage-I or in other words we choose a training checkpoint where vegetation detection is the best. This is important to preserve vegetation as in our next step we remove all pixels which are classified as background.

Finally, in stage-I, we simplify data by converting all detected background pixels to zeroes. This simplified data is passed through stage-II of semantic segmentation. This time semantic segmentation is applied with three output classes. Our neural network in stage-II learns better features to distinguish between crop, weed, and background. We have not developed any new neural network, instead, we have used standard neural networks as the backbone of our two-stage methodology. So, the novelty of our method lies in the effective usage of neural networks and simplification of the dataset by removing maximum pixels belonging to the background class in stage-I.

2.3. Backbone neural networks and segmentation models

UNet with Vanilla Mini CNN backbone is used in stage-I of our system to distinguish between vegetation and background, Vanilla Mini CNN is much smaller than the VGG network, and still, with experimentation, we have found that this UNet backbone can match the VGG16 backbone for vegetation detection in stage-I. In stage-II of the proposed system, we have used VGG16 as the backbone of UNet. Stage-II has a complex target as crop and weed have many similar characteristics so to efficiently learn discriminating features, a strong backbone with more neurons is required in stage-II. We have observed this phenomenon by experimenting with different lighter and heavier backbone neural networks. Different backbones can be utilized in both stages, we have achieved the best results with Vanilla Mini CNN backbone in stage-I and VGG16 backbone in stage-II. VGG small size 3 × 3 convolution kernel less computation power and its nonlinear optimization and strong fitting ability make it good to use as the backbone of the semantic segmentation network. The architecture of the proposed two-stage neural network is shown in Fig. 4.

2.4. Implementation

We have trained our 2-stage model from scratch and both stages are trained separately and independent of each other. We have made our trained classifier available (S. [19]) for researchers so that they could

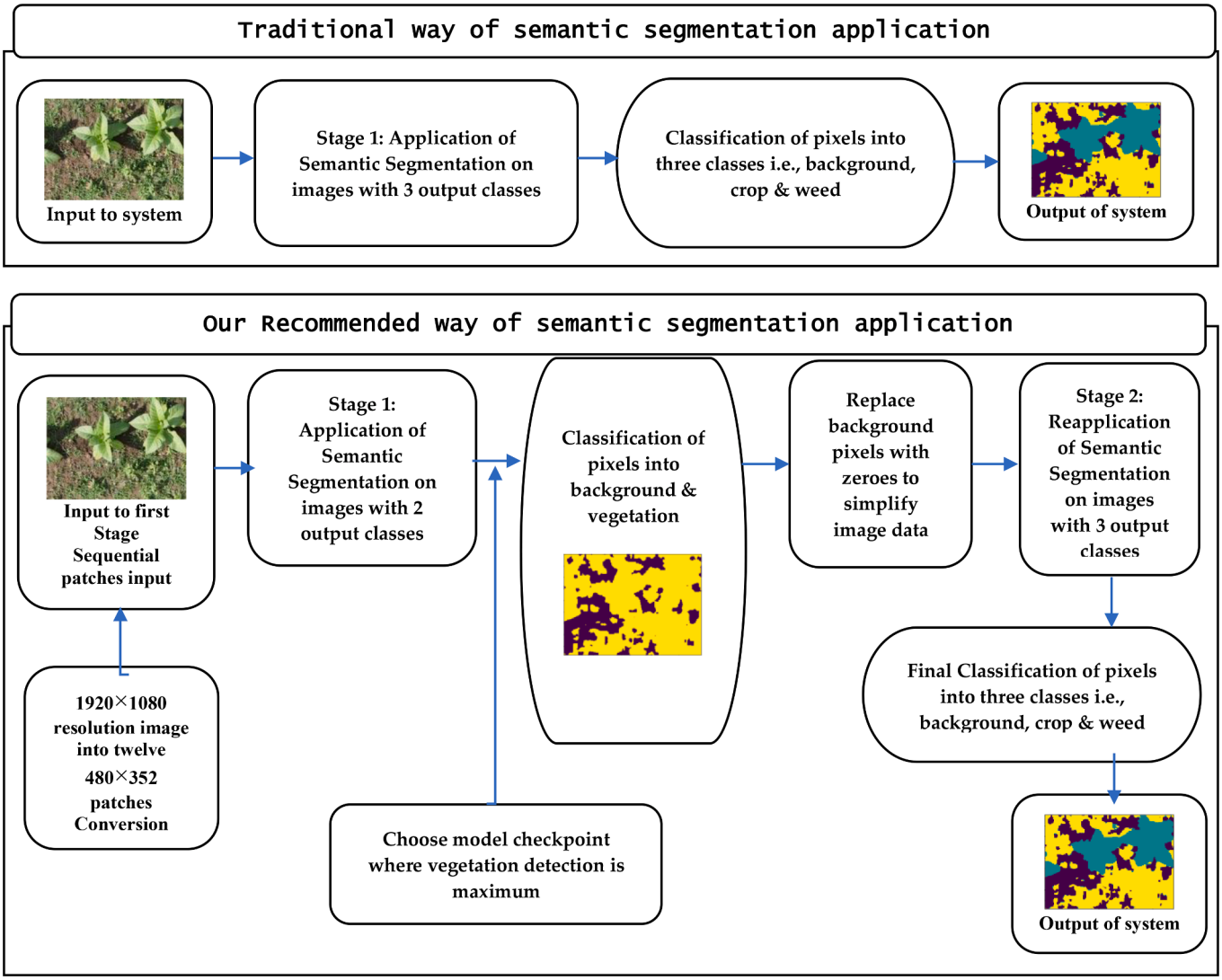

Fig. 2. Our proposed two-stage semantic segmentation method vs. traditional method of semantic segmentation.

use it directly or indirectly using transfer learning in tobacco weed classification applications. Keras with TensorFlow backend is used to apply deep learning methods. Our processing system specifications are eighth generation i5, RAM 16 GB with GTX 1050 NVIDIA GPU. The training is recorded with checkpoints. We employ a learning rate of 0.0001, which is typically regarded as suitable because it updates the weights minutely. Additionally, Adam and binary cross-entropy are chosen, respectively, as an optimizer and a loss function. We have used intersection over union (IOU) and mean intersection over union (MIOU) as an evaluation metrics. The dataset used in this research is high resolution, and at this resolution semantic segmentation cannot be applied due to hardware and software limitations, so we cropped non-overlapping 480 × 352 resolution patch images for training and testing.

3. Results and observations

In this section, results on different fields, comparative analysis with different approaches, and limitations of the algorithm are provided. Comparisons are made in the form of quantitative and qualitative results.

3.1. Results on seven different test fields

We have tested our 2-stage method with seven different tobacco field images, and we achieved MIOU ranging from 0.78 to 0.92 as shown in Table 2.

Visually our method has performed very well. In Appendix A seven images from seven different fields along with their output with our two-stage method is shown. These seven test images show different light conditions, shadows in a different orientation, different soil colors, soil dry and wet conditions, different tobacco growth stages and leaf sizes, and different weed types. Overall, a good tobacco and weed profile is received with our method.

3.2. Quantitative comparative results with traditional semantic segmentation application and our proposed method

Different segmentation models are used with different backbone models to see the efficiency. Table 3 shows quantitative results with semantic segmentation application when the traditional method is used, and it also compares results with our proposed two-stage application of semantic segmentation.

UNet has shown more accuracy than SegNet as a segmentation model. We have tried UNet with Vanilla Mini, VGG 16, and MobileNet backbones. In our experiments, Vanilla Mini performed best as the

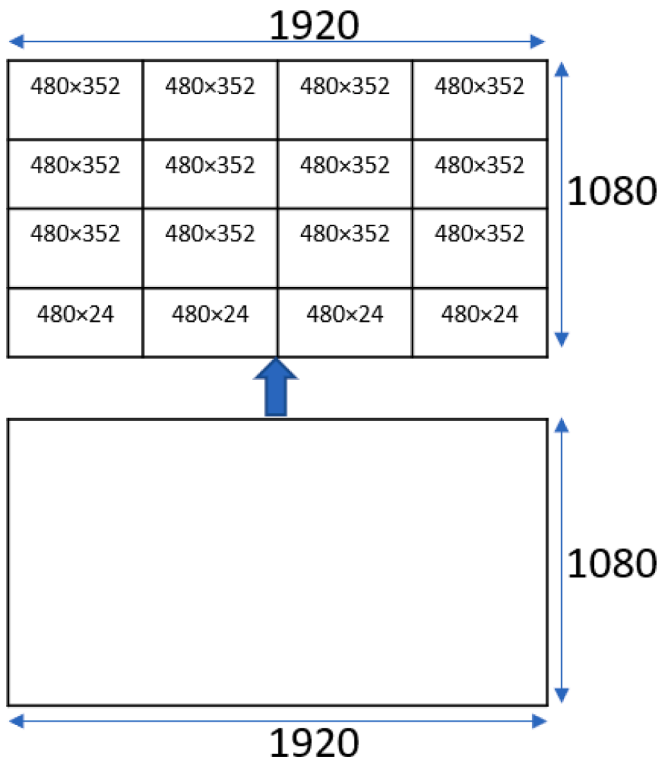

Fig. 3. 1080 P image patch conversion, 12 patches of size 480 × 352 size could be utilized from each original image.

backbone in stage-I, and in stage-II VGG16 performed the best as the backbone of the UNet segmentation network. Table 3, we have highlighted the best arrangements to use in stage-I and II of our system. When we compare the intersection over union IOU of crop = 0.85 and weed = 0.91 in stage-II using VGG16 as the backbone of UNet in Table 3, it has surpassed all other arrangements. This improvement in crop and weed detection is made possible due to the simplification of data at stage-I of our algorithm, where we converted most of the background pixels to zeroes. With the simplification of image data, in stage-II, the

Table 3
Tobacco Dataset Quantitative test Results—BG stands for the background, VEG stands for vegetation, C stands for Crop, W stands for weed, IOU stands for intersection over Union and MIOU stands for mean intersection over union.

Base Model/ Segmentation Model	Traditional 3-CLASS One Stage Semantic Segmentation Application 3-CLASS (BG+C + W)	Our proposed system application in two Stages Stage-I with Semantic Segmentation 2-CLASS (BG+VEG)	Stage-II with Semantic Segmentation 3-CLASS (BG+C + W)
Vanilla Mini CNN/ UNet	BG- IOU = 0.939 C- IOU = 0.709 W- IOU = 0.795 MIOU = 0.814	BG- IOU = 0.94 VEG-IOU = 0.89 MIOU = 0.91	BG- IOU = 0.949 C- IOU = 0.727 W- IOU = 0.8305 MIOU = 0.835
VGG 16/ UNet	BG- IOU = 0.92 C- IOU = 0.67 W- IOU = 0.77 MIOU = 0.79	BG- IOU = 0.93 VEG-IOU = 0.88 MIOU = 0.91	BG- IOU = 0.98 C- IOU = 0.85 W- IOU = 0.91 MIOU = 0.92
MobileNet/ UNet	BG- IOU = 0.913 C- IOU = 0.619 W- IOU = 0.740 MIOU = 0.757	BG- IOU = 0.912 VEG-IOU = 0.840 MIOU = 0.876	BG- IOU = 0.926 C- IOU = 0.637 W- IOU = 0.783 MIOU = 0.782
Vanilla CNN/ SegNet	BG- IOU = 0.912 C- IOU = 0.652 W- IOU = 0.750 MIOU = 0.771	BG- IOU = 0.912 VEG-IOU = 0.835 MIOU = 0.874	BG- IOU = 0.953 C- IOU = 0.748 W- IOU = 0.830 MIOU = 0.844
Resnet50/ SegNet	BG- IOU = 0.916 C- IOU = 0.66 W- IOU = 0.75 MIOU = 0.778	BG- IOU = 0.915 VEG-IOU = 0.843 MIOU = 0.879	BG- IOU = 0.965 C- IOU = 0.808 W- IOU = 0.883 MIOU = 0.885

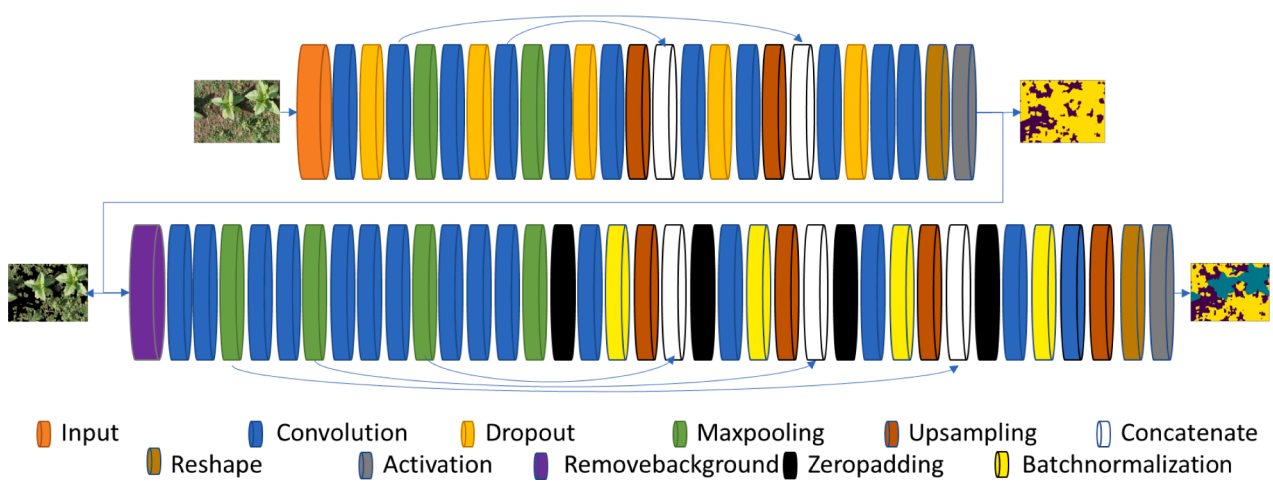

Fig. 4. Our proposed two-stage semantic segmentation neural network.

Table 2
Tobacco Dataset Quantitative test Results on Seven different fields—BG stands for the background, VEG stands for vegetation, C stands for Crop, W stands for weed, IOU stands for intersection over Union and MIOU stands for mean intersection over union.

Field no. / No. of tested images	119/ 936	120/ 120	133/ 120	134/ 120	154/ 120	163/ 120	171/ 120
Class wise and Mean Intersection over union	BG- IOU = 0.98 C- IOU = 0.85 W- IOU = 0.91 MIOU = 0.92	BG- IOU = 0.95 C- IOU = 0.57 W- IOU = 0.83 MIOU = 0.78	BG- IOU = 0.97 C- IOU = 0.76 W- IOU = 0.85 MIOU = 0.86	BG- IOU = 0.97 C- IOU = 0.81 W- IOU = 0.85 MIOU = 0.88	BG- IOU = 0.97 C- IOU = 0.85 W- IOU = 0.74 MIOU = 0.85	BG- IOU = 0.96 C- IOU = 0.72 W- IOU = 0.72 MIOU = 0.80	BG- IOU = 0.97 C- IOU = 0.77 W- IOU = 0.69 MIOU = 0.81

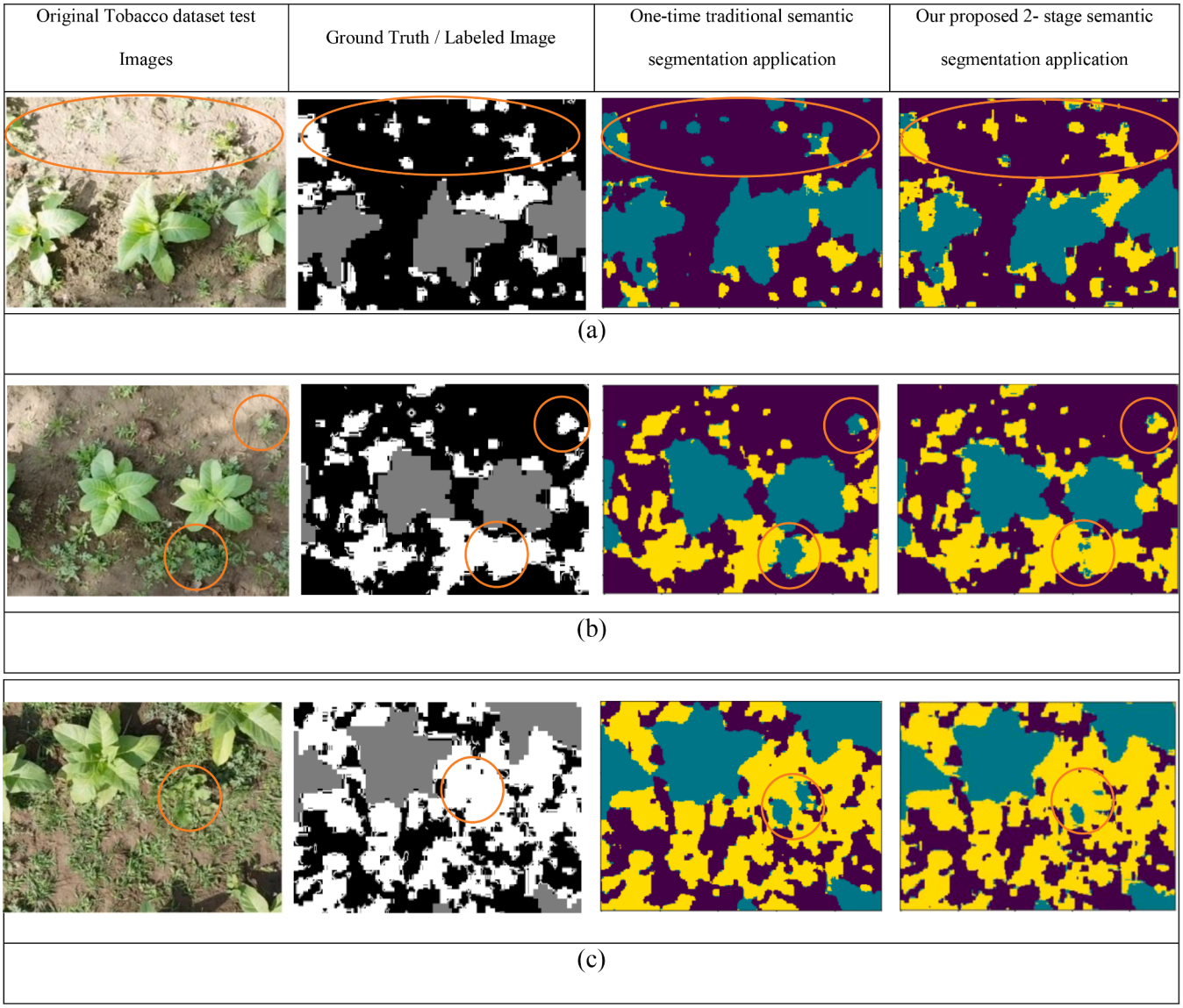

Fig. 5. Results with traditional one-stage semantic segmentation on the left, the original image in the middle, and our proposed 2-stage semantic segmentation on the same image on the right. In prediction results, cyan color shows predicted pixels as tobacco, yellow represents predicted weeds, and black shows predicted background.

algorithm learned more distinguishing parameters of crop and weed. We have also verified this phenomenon with visible results in Fig. 5. For the traditional single stage, three class semantic segmentation UNet with Vanilla Mini performed the best. In our proposed method, the same arrangement performed equally well compared to UNet with VGG16 backbone while detecting background and vegetation at stage-I of weed mapping. In stage-II, however, this arrangement failed to surpass UNet with the VGG16 backbone. In stage-II heavier VGG16 backbone has learned better crop and weed features.

Table 4
Tobacco Dataset Quantitative Comparative test Results—BG stands for the background, C stands for Crop, W stands for weed, IOU stands for intersection over Union, and MIOU stands for mean intersection over union.

	Traditional One Stage Semantic Segmentation Application [1] (I. [20])	Proposed two Stage system application (Ours)	
Base Model/ Segmentation Model	VGG 16/ SegNet	Stage I: VGG 16/ SegNet Stage II: VGG 16/ SegNet	Stage I: Vanilla Mini/UNet Stage II: VGG 16/ UNet
Classwise, IOU and MIOU	BG- IOU = 0.92 C- IOU = 0.67 W- IOU = 0.76 MIOU = 0.78	BG- IOU = 0.96 C- IOU = 0.81 W- IOU = 0.89 MIOU = 0.88	BG- IOU = 0.98 C- IOU = 0.85 W- IOU = 0.91 MIOU = 0.91
Total inference time to test field 119 data	~77 S	~154 S	~154 S
Inference time to test one 480 × 352 image	0.39 S	0.39 + 0.40 = 0.79 S	0.35 + 0.35 = 0.70 S

3.3. Visual comparative results with traditional semantic segmentation application and the proposed method

Results with traditional one-stage semantic segmentation, the original image, and our proposed 2-stage semantic segmentation are shown in Fig. 5. Visual results show an improved detection of weeds. In Fig. 5., we can visually confirm that our proposed two-stage application of semantic segmentation provides better classification and separability of tobacco and weed classes. In the first two images (Fig. 5. (a) and (b)), we can see challenging lighting conditions with direct sunlight in some areas and shade over others. Efficient detection of tobacco and weed in these images shows our proposed approach's robustness against variable lighting conditions. In Fig. 5. (a), we can see that traditional one-stage semantic segmentation has failed to predict some weeds in the blobs accurately; however, with simplified data and retraining, the proposed method has shown enhanced robustness in dealing with challenging lighting conditions. In Fig. 5. (b), we have highlighted two weed areas in the shade in which our proposed approach has shown better separability of classes. In Fig. 5. (c), we have highlighted a broadleaf plant in the image that is not tobacco but has a lot of similarity to tobacco than weeds. We have also highlighted its predicted output using traditional and our proposed method; the latter has shown better performance in the detection performance of weeds.

3.4. Comparative analysis with previous techniques

Table 4 compares us with [1] and (I. [20]). Both researchers have used SegNet architecture with a VGG16 backbone. First, we have implemented their technique, fine-tuned with a low learning rate of 0.0001 on our new tobacco crop dataset specifically, and then compared it with our two-stage implementation, where both stages are trained from scratch and then fine-tuned with the same learning rate of 0.0001.

We have shown a 2-stage semantic segmentation application with variations in base models and segmentation models. Our application of SegNet in both stages increased MIOU to 0.88 compared to 0.78 in one stage SegNet application of [1] and (I. [20]). We got the maximum performance of MIOU = 0.91 using Vanilla Mini and VGG 16 as backbones of UNet in stage-I and II, respectively. The inference time doubles with our approach, which indicates a limitation of our approach: our approach has almost double computational complexity. Conversely, we have seen a significant improvement in the crop, weed, and background detections. Our experimentation reveals that the UNet segmentation model is better than the SegNet model while applying it for crop and weed detection.

The results show that our two-stage application of semantic segmentation improves overall detections compared to the traditional semantic segmentation application.

We have checked the inference time for one image of size 480 × 352; it is below one second, as shown in Table 4. The inference time is almost double for our 2-stage approach; it is still appropriate if the processing of images is done offline and on a higher end processing unit.

We observed that our proposed 2-stage approach gave better results than traditional one-stage deep learning applications with the same neural network usage. Our proposed system can be used in selective spraying tobacco and its weeds. Selective spraying on industrial tobacco crops would save agrochemicals, saving expenditure on agrochemicals, and high yield with less expense would increase tobacco profit. Selective targeted spray on tobacco crops and their weeds would also decrease soil pollution, a major health and environmental concern.

3.5. Comparison with two-stage weed classifier network

This section elaborates on the comparison between MTS-CNN by Kim Y. H. and Park K. R. [10], (which contain two UNets connected in series) and our proposed two stage method. For this experimentation input image sizes are kept same i.e., 480 × 352 for both MTS-CNN and our proposed model which is good for the comparison of results. Encoder-decoder sizes for MTS-CNN are kept three in both stages of UNets in our experimentation. First stage of our model uses encoder size of two and second stage uses encoder size of three. Implementation hyperparameters are kept same for MTS-CNN and our proposed model as described in Section 2.4. Fig. 6 shows Comparison of Proposed Model with MTS-CNN.

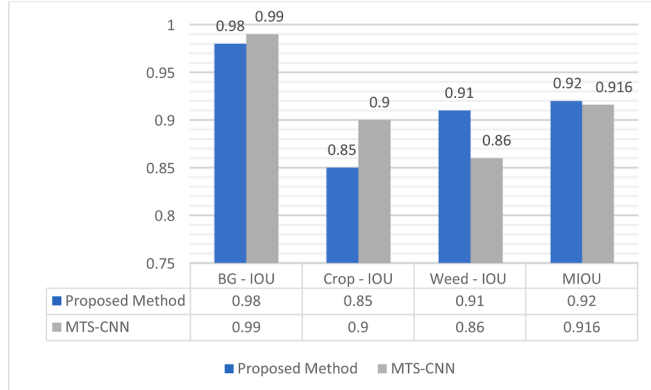

Fig. 6. Comparison of Proposed Model with MTS-CNN, showing class wise intersection over union (IOU) and mean intersection over union (MIOU), BG stands for background.

The results with both methods are comparable, although better crop pixel prediction is seen in MTS-CNN prediction and better weed classification is seen with our proposed method the overall MIOU is comparable.

The computational complexity of both models for 480 × 352 size input images, we see that our proposed model (471,586 + 12,321,603 = 12,793,189 trainable parameters) is much smarter than MTS-CNN (12,321,603 + 12,321,603 = 24,643,206 trainable parameters). our proposed two-stage model could be thought of as the optimized version of MTS-CNN, with half computational complexity.

3.6. Limitation of our proposed method

The limitation of the proposed method is that there are few crop and weed pixels in stage-I which are wrongly predicted as background, and those pixels are removed when we simplified the image by removing the background. However, those pixels are low in numbers; so overall we have achieved a better separability of crop and weed at the expense of a few missed crop and weed pixels detections.

To limit the classification of crop and weed pixels into the background in stage-I, we have selected a checkpoint of the neural network in stage-I so that fewer crop and weed pixels get misclassified as background. This step is important because all the pixels which are misclassified as background in stage-I will be removed. The application of semantic segmentation in two stages has increased the detection accuracy but at the cost of computational complexity. We have observed that in stage-I shallower and smaller semantic segmentation model is required as compared to stage-II where a segmentation network with more neurons serves the purpose of good detection. This observation is justifiable as we know that the background-vegetation detection problem is simpler as compared to the crop-weed detection problem. By applying semantic segmentation in two stages, we observed the increase of IOU (intersection over union) of crop and weed as compared to its application one time. To reduce the complexity of this approach for now we have used a smaller backbone i.e., vanilla mini in stage-I.

4. Conclusion & future work

This paper proposes a two-stage application of semantic

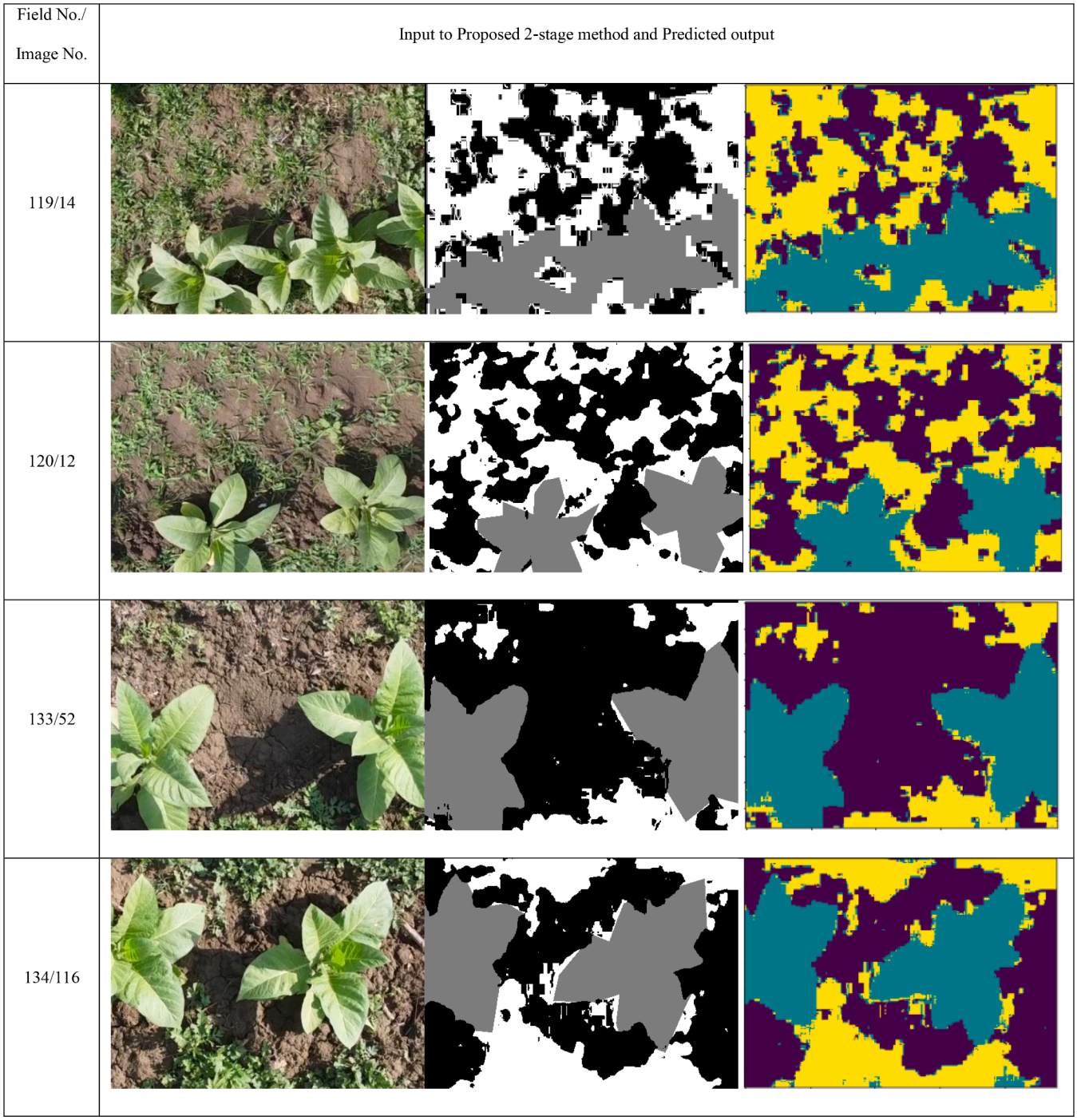

Fig. A1. Predicted results with our proposed 2-stage semantic segmentation on different images from seven various fields. (In prediction results, cyan color shows predicted pixels as tobacco, yellow represents predicted weeds, and black shows predicted background.).

segmentation; in the first stage of semantic segmentation dataset is simplified by detecting background and vegetation, and then in the second stage of semantic segmentation application, better classification of crop and weed is achieved. The first contribution of the paper is a new labeled aerial tobacco dataset that is publicly available for the research community for autonomous weed removal, spray application, and yield estimation in tobacco fields. The crop and weed detection methodology described in this research could be used to spray tobacco crops efficiently.

Our experiment shows that a simpler semantic segmentation model can perform efficiently in the first stage. However, in the second stage, a deeper model is needed to separate crops and weeds efficiently. We have compared existing techniques and found that our 2-stage approach was performing better than existing single-stage deep learning techniques under identical neural network backbones. We achieved satisfactory results for tobacco-weed classification especially. A Trained classifier is uploaded online, and it could be used directly or indirectly using transfer learning for background/tobacco/weed classification applications. Application of semantic segmentation with our proposed method gives a satisfactory separation of all three classes under complex lighting conditions; at the increased computational cost compared to existing implementations, computational complexity with our approach was

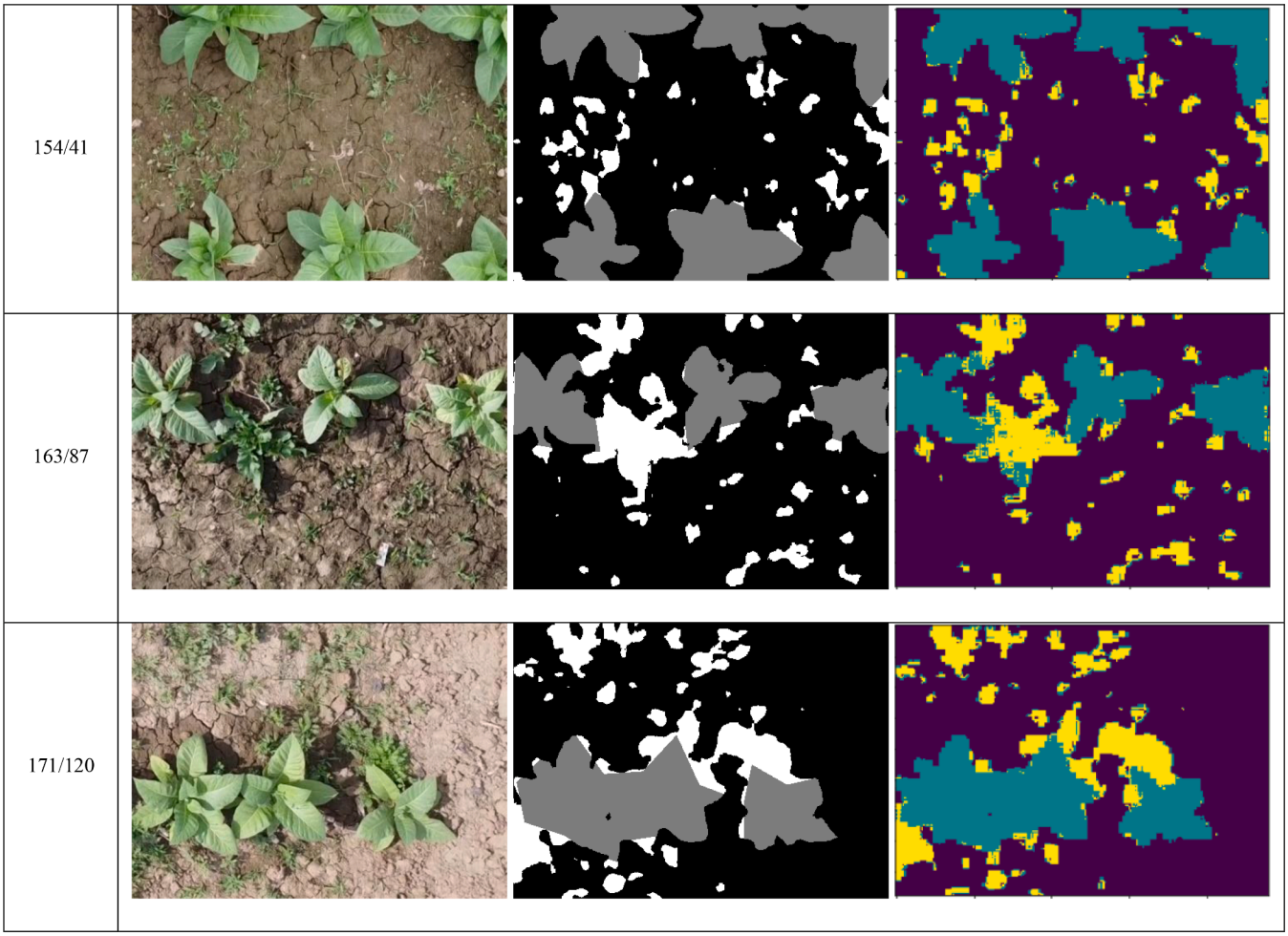

Fig. A1. (*continued*).

double; however, modern embedded systems available in the market can handle this complexity, inference time with our approach was suitable to go for real-time application, and our proposed model could be optimized further in future. Our future work would aim to optimize the proposed two-stage methodology and implement our proposed model on NVIDIA Jetson Nano board. Also, we are planning to implement our approach on more commercial and food crops to see its generalization ability.

Data Availability

Data will be made available on request.

Acknowledgements

This work is funded by the Higher Education Commission of Pakistan and the National center for Robotics and Automation (DF-1009–31). We thank Pakistan Tobacco Company for helping us find farms for data collection.

Appendix A

Fig. A1

References

[1] 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), https://doi.org/10.1016/j.compag.2019.105091.

[2] M. Alam, M.S. A, M. R, M. T, M.U. K, M.T K, Real-time machine learning based crop/weed detection and classification for variable-rate spraying in precision agriculture, in: International Conference on Electrical and Electronics Engineering, 2020, pp. 273–280.

[3] C.C. Andrea, B. Mauricio Daniel, J.B. Jose Misael, Precise weed and maize classification through convolutional neuronal networks, in: *2017 IEEE 2nd Ecuador Technical Chapters Meeting, ETCM 2017, 2017-Janua*, 2018, pp. 1–6, https://doi.org/10.1109/ETCM.2017.8247469.

[4] B. Espejo-Garcia, N. Mylonas, L. Athanasakos, S. Fountas, I. Vasilakoglou, Towards weeds identification assistance through transfer learning, Computers and Electronics in Agriculture 171 (2020), https://doi.org/10.1016/j.compag.2020.105306.

[5] J.S. Holt, Principles of Weed Management in Agroecosystems and Wildlands, in: Invasive Weed Symposium 18, 2004. https://www.jstor.org/stable/3989691?seq=1&cid=pdf.

[6] N. Iqbal, S. Manalil, B.S. Chauhan, S.W. Adkins, Investigation of alternate herbicides for effective weed management in glyphosate-tolerant cotton, Archives of Agronomy and Soil Science 65 (13) (2019) 1885–1899, https://doi.org/10.1080/03650340.2019.1579904.

[7] A.J. Ishak, S.S. Mokri, M.M. Mustafa, A. Hussain, Weed detection utilizing quadratic polynomial and ROI techniques, in: *2007 5th Student Conference on Research and Development, SCORED, December*, 2007, pp. 0–4, https://doi.org/10.1109/SCORED.2007.4451360.

[8] Y. Jiang, C. Li, A.H. Paterson, J.S. Robertson, DeepSeedling: deep convolutional network and Kalman filter for plant seedling detection and counting in the field, Plant Methods 15 (1) (2019), https://doi.org/10.1186/s13007-019-0528-3.

[9] Y. Karimi, S.O. Prasher, R.M. Patel, S.H. Kim, Application of support vector machine technology for weed and nitrogen stress detection in corn, Computers and Electronics in Agriculture 51 (1–2) (2006) 99–109, https://doi.org/10.1016/j.compag.2005.12.001.

[10] Y.H. Kim, K.R. Park, MTS-CNN: multi-task semantic segmentation-convolutional neural network for detecting crops and weeds, Computers and Electronics in Agriculture 199 (August 2022) (2022), 107146.

[11] F.J. Knoll, V. Czymmek, L.O. Harders, S. Hussmann, Real-time classification of weeds in organic carrot production using deep learning algorithms, Computers and Electronics in Agriculture 167 (2019), https://doi.org/10.1016/j.compag.2019.105097.

[12] P. Lameski, E. Z, A. K, Review of automated weed control approaches: an environmental impact perspective, in: International Conference on Telecommunications, 2018, pp. 132–147.

[13] V.N.T. Le, S. Ahderom, K. Alameh, Performances of the lbp based algorithm over cnn models for detecting crops and weeds with similar morphologies, Sensors (Switzerland) 20 (8) (2020), https://doi.org/10.3390/s20082193.

[14] A. Milioto, P. Lottes, C. Stachniss, REAL-TIME BLOB-WISE SUGAR BEETS VS WEEDS CLASSIFICATION for MONITORING FIELDS USING CONVOLUTIONAL NEURAL NETWORKS, in: ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences 4, 2017, pp. 41–48, https://doi.org/10.5194/isprs-annals-IV-2-W3-41-2017.

[15] Nkemelu, D.K., Omeiza, D., & Lubalo, N. (2018). *Deep Convolutional Neural Network for Plant Seedlings Classification*. http://arxiv.org/abs/1811.08404.

[16] V. Partel, S. Charan Kakarla, Y. Ampatzidis, Development and evaluation of a low-cost and smart technology for precision weed management utilizing artificial intelligence, Computers and Electronics in Agriculture 157 (December 2018) (2019) 339–350, https://doi.org/10.1016/j.compag.2018.12.048.

[17] Ruckelshausen, A., Biber, P., Dorna, M., Gremmes, H., Klose, R., Linz, A., Rahe, R., Resch, R., Thiel, M., Trautz, D., & Weiss, U. (2009).

[18] S. Imran Moazzam, U. S. K. T. N. W. S. Q. F. K. (2022a). *Tobacco Dataset*. https://data.mendeley.com/v1/datasets/5dpc5gbgpz/draft?a=888c40e2-c349-4c1a-ab69-0f478c6980c5,Https://1drv.Ms/u/s!Ao5jMGloq7Xlg70uK9br2hRbNvtClA?E=PfrWFq.

[19] S. Imran Moazzam, U. S. K. T. N. W. S. Q. F. K. (2022b). *Tobacco two-stage Classifiers*. Https://1drv.Ms/u/s!Ao5jMGloq7Xlg79dHZX7tU3e9POB3w?E=3NmilR.

[20] I. Sa, M. Popović, R. Khanna, Z. Chen, P. Lottes, F. Liebisch, J. Nieto, C. Stachniss, A. Walter, R. Siegwart, WeedMap: a large-scale semantic weed mapping framework using aerial multispectral imaging and deep neural network for precision farming, Remote Sens (Basel) 10 (9) (2018), https://doi.org/10.3390/rs10091423.

[21] S. Sabzi, Y. A.-G, J.I. A, An automatic visible-range video weed detection, segmentation and classification prototype in potato field, Heliyon 6 (5) (2020).

[22] A. Wendel, J. Underwood, Self-supervised weed detection in vegetable crops using ground based hyperspectral imaging, in: *Proceedings - IEEE International Conference on Robotics and Automation, 2016-June*, 2016, pp. 5128–5135, https://doi.org/10.1109/ICRA.2016.7487717.