Weed Detection among Crops by Convolutional Neural Networks with Sliding Windows

Karthik Kantipudi[a], Chih Lai[a], Cheol-Hong Min[b], Ron C. Chiang[a]

[a] Graduate Programs in Software, University of St. Thomas, St Paul, Minnesota

[b] School of Engineering, University of St. Thomas, St Paul, Minnesota

A paper from the Proceedings of the
14th International Conference on Precision Agriculture
June 24 – June 27, 2018
Montreal, Quebec, Canada

Abstract.

One of the primary objectives in the field of precision agriculture is weed detection. Detecting and expunging weeds in the initial stages of crop growth with deep learning technique can minimize the usage of herbicides and maximize the crop yield for the farmers. This paper proposes a sliding window approach for the detection of weed regions using convolutional neural networks. The proposed approach involves two processes: (1) Image extraction and labelling, (2) building and training our neural network. In the image extraction process, sub-images are extracted by slicing the original images into sub-images. Subsequently each sub-image is labelled based on the given annotation images. In the next process for building the network architecture, convolutional neural network model is implemented with 20 layers consisting of one image input layer, four 2-D convolutional layers, six rectified linear unit (ReLU) layers, four 2-D max pooling layers, three fully connected layers, one softmax layer, and one final classification layer. Various collections of sub-images gathered by various sliding window sizes were passed into this network to determine the best sliding window size that resulted in higher true weed detection rate and lower percentage of crop wastage. After the ratios between true weed detection rate and crop wastage values were computed for each sliding window sizes, it was found out that the sliding window size of [80 80] resulted in the maximum ratio with true weed detection rate and crop wastage values of 63.28% and 13.33% respectively. These findings reveal that the sliding window size of [80 80] was able to predict the true weed regions in the crops field imagery with 63.28% accuracy that causes least damage to the crops due to predicting crops as weeds.

Keywords. *Convolutional Neural Network, True weed detection rate, Crop wastage, sliding window, sub-image*

1. Introduction

Weeds are unwanted plants which not only hamper crop growth but also damage the crop yield significantly, thereby causing great losses to the farmers economically. Various technologies have been developed in the field of precision agriculture to help farmers detect and eradicate weeds in the early stages of crop growth. These technologies help farmers to optimize the pesticides and herbicides input for effective removal of weeds from crop field, eventually leading to better crop yield.

1.1. Related Work

Initial efforts were put to detect weeds from crops by automatic computer vision-based approach where image segmentation and Bayesian decision making strategies were involved when weeds appear irregularly distributed within the crop's field [1]. Leaf and canopy spectral relative reflectance values were gathered with the help of field spectroscopy tools to detect grasses and broadleaf weeds among crops [2]. In Ref. [3] a membership function on fuzzy logic approach was developed to classify vegetable crops and weed species using morphological and color features of digital images.

In Ref. [4] a CNN (Convolutional Neural Network) was proposed for real-time semantic segmentation of weeds and crops which exploits existing vegetation indexes. Related research work also focuses on using the recently developed encoder-decoder cascaded CNN, SegNet to infer dense semantic classes using multi spectral images collected by micro aerial vehicle (MAV) [5]. Ref [6] exploits two CNNs, a light-weight CNN to extract the pixels that represent projections of 3D points of the green vegetation and a deeper CNN that was used to classify the extracted pixels between the crop and weed classes. Ref. [7] utilizes Fast Fourier Transform and leaf edge density to classify weeds from corn field in real-time. Ref. [8] uses two approaches, one is Bayes classifier and the other corresponds to the application of morphological operators on weed and crop areas in edge images for detection of weeds in the lawns.

1.2. Dataset

The dataset consists of 60 images along with annotations available online (https://github.com/cwfid/dataset). These images were acquired by two German researchers Sebastian Haug and Jörn Ostermann with the help of autonomous field robot Bonirob in an organic carrot farm when the plants were still at their early growth stage [9]. Also, at the time when the images were captured for the data collection, both the weeds and crops were of approximately same size.

2. Approach:

The weed detection process consists of two sub-processes: Image extraction with labelling the input and building the network architecture. The image extraction process divides each incoming training image into sub-images where the collections of these sub-images are sent to the Convolutional Neural Network models to predict the potential weed regions in the test images.

2.1. Image extraction and labelling the input

After the division of dataset into training and testing sets in 2:1 ratio, sliding windows of various sizes are scanned over each training image to slice them into sub-images. If $w \times w$ is the sliding window size that is scanned over each image of size $p \times q$ with no padding and stride 'w', then the number of sub-images(N) formed for an image for that window size would be

$$N = (p*q) / (w*w) \quad (1)$$

Each collection consists of sub-images of same size for a given sliding window size rolled over on each training image.

After obtaining the training set consisting of sub-images of same size, labelling of the input sub-images is also done to pass the data set into the convolutional neural network models for predicting weed, crop and soil regions in the testing set. As each sub-image in the collection contains any number of weed, crop and soil pixels, labelling of these images is done by utilizing the annotation images from original dataset.

For this purpose, the same sliding windows which are scanned on each of the original images are simultaneously scanned on the annotation images too, to obtain the labels for the corresponding sub-image in an image. The labels of the sub-images are given as '0','1','2' for soil, weed and crop categories respectively. These labels are allotted to each sub-image based on the percentages obtained by calculating the number of pixels for each category in the annotation sub-image divided by the total number of pixels in the same annotation sub-image. After calculating the percentages per each category in the sub-image, the label for a sub-image is decided based upon the category which had the maximum percentage pixel share in the corresponding annotation sub-image.

In the entire process of our research, MATLAB (Matrix Laboratory) software was used as a platform for the experiments that were conducted on.

The below MATLAB code shows the key steps followed for the assignment of the labels to the sub-images.

MATLAB code:

```
CropPercent = Num_CropPixels / numel(sub_image_Annotation);
WeedPercent = Num_WeedPixels / numel(sub_image_Annotation);
SoilPercent = Num_SoilPixels / numel(sub_image_Annotation);
if ((SoilPercent > WeedPercent) && (SoilPercent > CropPercent))
        Labels_sub_image = 0;
elseif ((WeedPercent > SoilPercent) && (WeedPercent > CropPercent))
        Labels_sub_image = 1;
else
        Labels_sub_image = 2;
end
```

Here,

sub_image_Annotation = sliced sub-image of a given annotation image

Num_CropPixels = Number of Crop pixels in the image "*sub_image_Annotation*"

Num_WeedPixels = Number of Weed pixels in the image "*sub_image_Annotation*"

Num_SoilPixels = Number of Soil pixels in the image "*sub_image_Annotation*"

CropPercent = Percentage of Crop pixels in the image "*sub_image_Annotation*"

WeedPercent = Percentage of Weed pixels in the image "*sub_image_Annotation*"

SoilPercent = Percentage of Soil pixels in the image "*sub_image_Annotation*"

Labels_sub_image = Label for the sliced sub-image

numel function gives total number of elements present in the matrix

2.2. Network architecture

After labels were obtained for their respective sub-images from the annotation images, these sub-images and their labels are passed into the convolutional neural network models to predict the potential weed regions in the initially partitioned testing set.

The model architecture consists of one image input layer, four 2-D convolutional layers, six rectified linear unit (ReLU) layers, four 2-D max pooling layers, three fully connected layers, one softmax layer, and one final classification layer.

The sequence of layers of the model architecture is shown in Fig. 1. The input for image input layer is nothing but the sliced sub-image size when a given size of sliding window is scanned on original training image. The second layer in the sequence is the Convolutional 2D Layer which takes in the inputs of filter size as [11 11] and 96 filters with a padding size of two extra borders. The third layer is the ReLU layer which is used to improve the network by speeding up the training process. Maxpooling 2D layer is the fourth layer in the network architecture which performs down-sampling with a pooling size of 3*3 and a stride of two pixels. The input for this layer is nothing but the output produced by the first ReLU Layer. The second convolutional 2D layer (5th layer) produces the convolutions with a filter size of [5 5] and 256 filters with the extra borders of 2 rows and 2 columns. The third and fourth Convolutional 2D Layers (8th layer and 11th layers respectively) take in the inputs of filter sizes as [5 5] and the number of filters for both the layers being 384 and 256 respectively. The input hyper parameters remain the same for all max pooling 2d layers in the network. Two fully connected layers with 4096 neurons and one another final fully connected layer with 3 neurons were used in the final layers to flatten the incoming data and for classifying the input image into 3 classes.

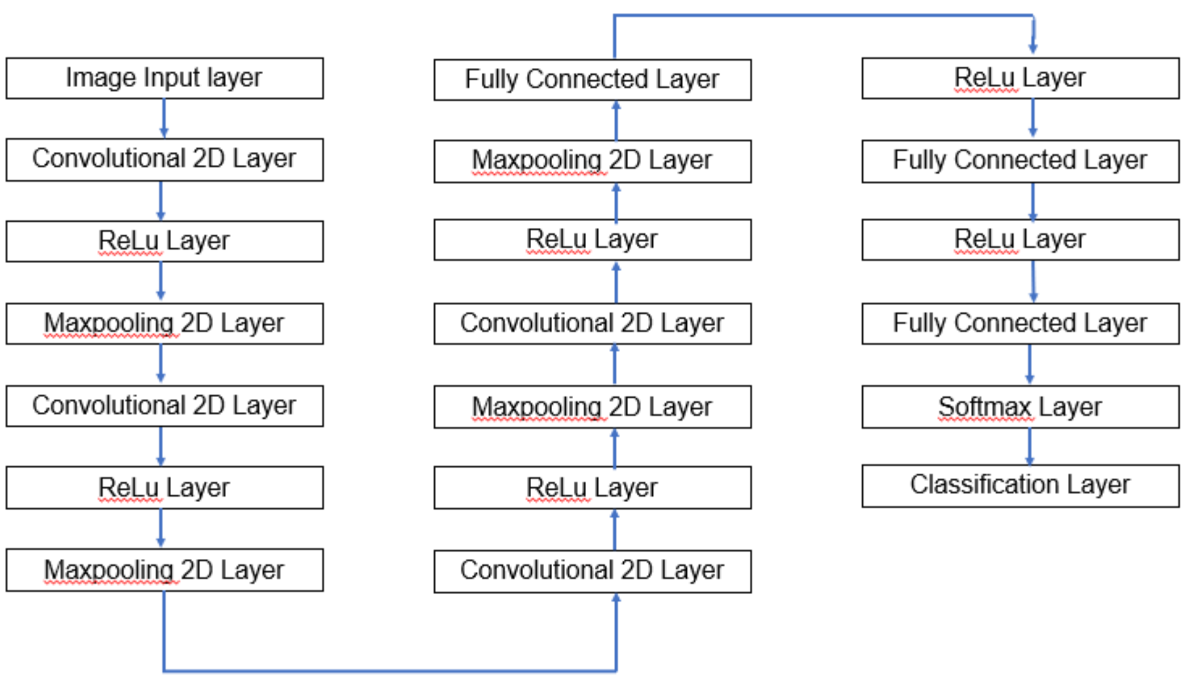

Fig. 1 shows the sequence of different layers in Convolutional Neural Network model that was used for predictions

3. Experimental evaluation:

3.1. Performance measures

After building the model, the sub-image collection for a sliding window size along with labels is passed into the network to predict the potential weed regions for measuring the performances of the model in terms of true weed detection rate and crop wastage.

Here,

$$WD = 2^* \text{ Precision}^* \text{ Recall} / (\text{Precision} + \text{Recall}) \quad (2)$$

where,

WD = True weed detection rate.

Precision = mean (No. of ground truth boxes of weeds that were able to touch (1 pixel) at-least one prediction box of weeds / No. of ground truth boxes of weeds) of all testing images.

Recall = mean (No. of prediction boxes of weeds that were able to touch at-least one ground truth boxes of weeds / No. of prediction boxes of weeds) of all testing images.

Also, crop wastage(CW) is defined as

$$CW = \text{mean (Percentage of crop pixels contained in the prediction boxes of weeds) of all images} \quad (3)$$

The collections of sub-images sliced by various sizes of sliding windows are passed into the same network to choose the optimum sliding window size that results in high true weed detection rate and lower crop wastage as well.

Table 1. shows recall, precision, WD (F_score or Weed Detection rate), CW (Crop Wastage) and Ratio (WD/CW) values for different sliding windows when various collections of sub-images were passed into the CNN model.

Window size	Recall	Precision	WD	CW	Ratio (WD/CW)
[30 30]	0.9114	0.7825	0.8266	0.7123	1.1605
[35 35]	0.8053	0.7926	0.7901	0.6213	1.2717
[45 45]	0.7477	0.8364	0.7781	0.3971	1.9595
[55 55]	0.7530	0.8441	0.7799	0.3949	1.9749
[60 60]	0.755	0.8582	0.7796	0.4771	1.6340
[65 65]	0.5753	0.8231	0.6590	0.2477	2.6604
[70 70]	0.6034	0.8298	0.6876	0.3240	2.1222
[75 75]	0.5330	0.8617	0.6292	0.2410	2.6108
*[80 80]	0.5075	0.9303	0.6328	0.1333	4.7472
[85 85]	0.5420	0.9288	0.6507	0.2180	2.9849
[90 90]	0.3940	0.7306	0.4972	0.1248	3.9840
[95 95]	0.3662	0.8254	0.4864	0.1509	3.2233
[100 100]	0.4138	0.7554	0.5076	0.1514	3.3527
[105 105]	0.3654	0.7602	0.4674	0.1672	2.7954
[110 110]	0.2622	0.6667	0.3665	0.0920	3.9837
[115 115]	0.2635	0.6905	0.3645	0.1461	2.4949
[120 120]	0.2351	0.6667	0.3271	0.0925	3.5362

*Best Sliding window size = [80 80] with ratio = 4.7472

3.2. Best sliding window size

To decide upon the sliding window size that results in high true weed detection rate and low crop wastage, ratio based upon WD and CW was calculated for all the sliding window results. Best sliding window size is that size which produced highest ratio among all the ratios that were calculated for all the sliding window sizes.

From the above shown results in Tab 1., best sliding window size was identified to be [80 80] with the ratio of 4.7492 where WD and CW values where found out to be 0.6328 and 0.1333 respectively.

From Fig.2, the color variations among various measures show that the smaller window sizes tend to result in a higher true weed detection rate and higher crop wastage values whereas the larger window sizes tend to produce lower true weed detection rates and lower crop wastage values of which both extremes are undesirable.

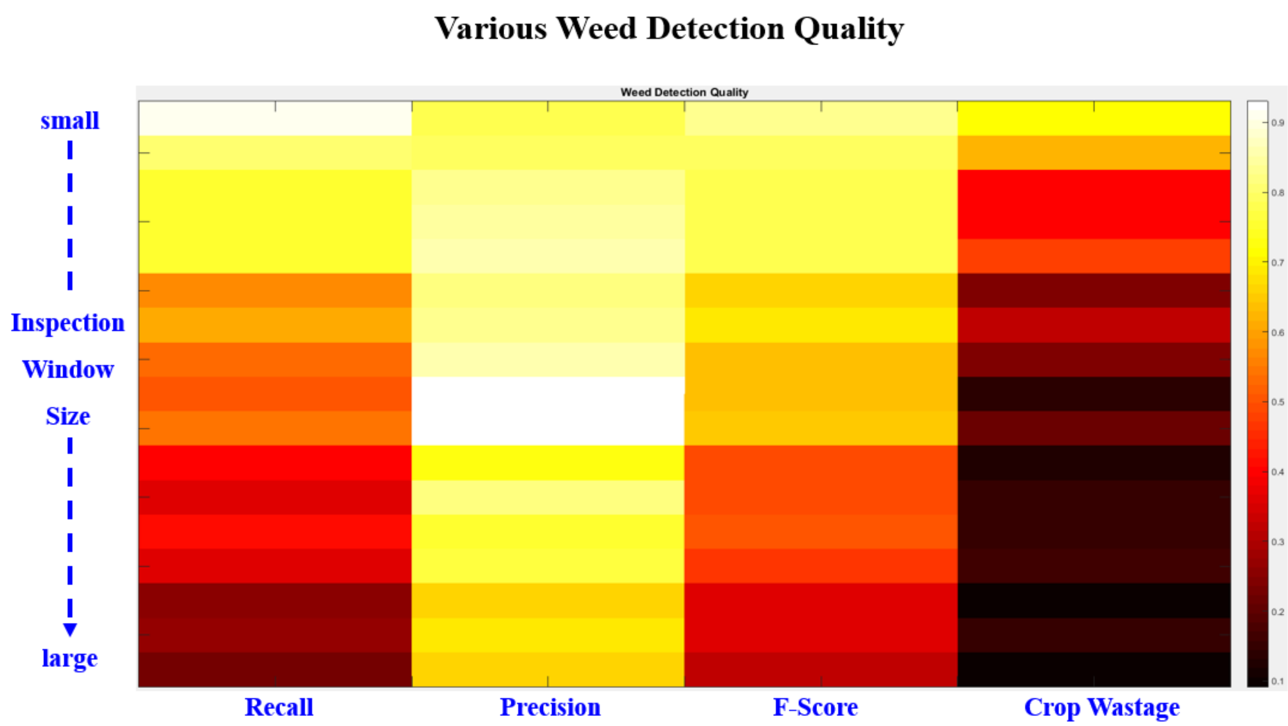

Fig. 2 shows the color variations of Recall, Precision, F-Score and Crop Wastage values with increasing window sizes

4. Conclusion:

A sliding window approach was proposed to predict the most potential weed regions in the organic carrot farm imagery with the likelihood of causing least damage to the crops. The experiments conducted showed that varying window sizes impact the weed detection rate and percentage of crop wastage values enormously. When the sliding window sizes are too small, it is found that although true weed detection rates are very high, the crop damage rates are also high as well. This indicates that even though small sliding window sizes are good predictors for detection of true weed regions, they are not good predictors for the prediction of crop pixels and there is a high chance of farmers spraying their chemicals on crop regions too, thereby damaging the crop health. When the sliding windows sizes are too large, even though crop wastage percentage is too small when weed regions were predicted, the true weed detection rate is not significant enough to detect all the weeds in the farm. So, when a ratio was calculated between true weed detection rate and crop wastage, it was found that sliding window size of [80 80] would result in an effective detection of weeds with 63.28% while causing the least damage to the crop with 13.33%.

Acknowledgements

This work was supported by grant from the Precision Agriculture research group from the University of St Thomas.

5. References

Article by DOI

[1] Tellaechea, A., Burgos-Artizzub, X., P., Pajaresa, G., Ribeirob, A. (2007). A vision-based method for weeds identification through the Bayesian decision theory. The Journal of the Pattern Recognition Society, 41. 521-530. doi: 10.1016/j.patcog.2007.07.007

Proceedings paper

[2] Shapira, U, Karnieli, A., Bonfil, D. (2010). Weeds detection by ground-level hyperspectral data. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives. 38.

Article by DOI

[3] Hemming, J., Rath, T. (2001). PA—Precision Agriculture: Computer-Vision-based Weed Identification under Field Conditions using Controlled Lighting. Journal of Agricultural Engineering Research. 78. 233-243. doi: 10.1006/jaer.2000.0639.

Online document

[4] Milioto, A., Lottes, P., Stachniss, C. (2017). Real-time Semantic Segmentation of Crop and Weed for Precision Agriculture Robots Leveraging Background Knowledge in CNNs. https://arxiv.org/abs/1709.06764.

Article by DOI

[5] Sa, I., Chen, Z., Popovic, M., Khanna, R., Liebisch, F., Nieto, J., Siegwart, R. (2017). weedNet: Dense Semantic Weed Classification Using Multispectral Images and MAV for Smart Farming. IEEE Robotics and Automation Letters. PP. doi:10.1109/LRA.2017.2774979.

Proceedings paper

[6] Ciro, P., Daniele, N., Alberto, P. (2017). Fast and Accurate Crop and Weed Identification with Summarized Train Sets for Precision Agriculture. Intelligent Autonomous Systems 14: Proceedings of the 14th International Conference IAS-14 (pp.105-121).doi 10.1007/978-3-319-48036-7_9

Article by DOI

[7] Nejati, H., Azimifar, Z., Zamani, M. (2008). Using fast fourier transform for weed detection in corn fields. 1215 - 1219. doi:10.1109/ICSMC.2008.4811448.

Article by DOI

[8] Watchareeruetai, U., Takeuchi, Y., Matsumoto, T., Kudo, H., Ohnishi, N. (2006). Computer Vision Based Methods for Detecting Weeds in Lawns. Machine Vision and Applications 17. 287-296. doi:10.1109/ICCIS.2006.252275.

Proceedings paper

[9] Haug, S., Ostermann, J. (2014).A Crop/Weed Field Image Dataset for the Evaluation of Computer Vision Based Precision Agriculture Tasks. Computer Vision - ECCV 2014 Workshops: Zurich, Switzerland, September 6-7 and 12, 2014, Proceedings, Part IV (pp.105-116). doi 10.1007/978-3-319-16220-1_8