# WEED CLASSIFICATION IN HYPERSPECTRAL REMOTE SENSING IMAGES VIA DEEP CONVOLUTIONAL NEURAL NETWORK

*Adnan Farooq, Jiankun Hu, Xiuping Jia*

School of Engineering and Information Technology
University of New South Wales
Canberra, ACT 2600, Australia

## ABSTRACT

Automatic weed detection and mapping are critical for site-specific weed control in order to reduce the cost of farming as well as the impact of herbicides on human health. In this paper, we investigate patch-based weed identification using hyperspectral images. Convolutional Neural Network (CNN) is evaluated and compared with the Histogram of Oriented Gradients (HoG) for this purpose. Suitable patch sizes are investigated. The limitation of RGB imagery is demonstrated. The experimental results indicate that the overall accuracy of the weed classification using CNN increases with the increasing number of bands used. With more bands, CNN extracts more powerful and discriminative features and leads to improved classification as compared to the traditional HoG feature extraction method. The computational load of CNN, however, is slightly increased with the increasing number of bands.

***Index Terms*—** Hyperspectral images, weed mapping, Histogram of Oriented Gradients (HoG), Convolutional Neural Network (CNN)

## 1. INTRODUCTION

To control the undesired weed species while promoting the development of valuable crops is important and, nowadays, it is expected to achieve the task effectively. Currently, weed control system mainly relies on chemical herbicides over the complete agriculture regions. This process is costly, as the farmers spend thousands of dollars annually on herbicides for commercial farming. Apart from the financial costs, there are severe concerns about the impact of using chemical herbicides on human and the health of desired crops. Furthermore, it reduces the overall production of crops by 10% per year [1]. Therefore, it is important to detect undesired weed species and limit the use of chemical herbicides.

With the advancement in remote sensing, many satellites and aerial sensors provide imagery with high spatial resolution, wide coverage and regular observations. Weed detection from the hyperspectral images, however, is a challenging task due to the large intraclass variation of weeds, including sizes and conditions, and similar colour to other vegetation and weeds. The detection is more difficult if there are occlusion with crops, complex background environment (i.e., soil), changing illumination, and shadowing effect, etc. To address these challenges, extensive research has been conducted [1].

Two different types of approaches have been developed by the remote sensing community for classification tasks, i.e., pixel-based and object-based [2] [3] [4] [5]. In recent years, object-based approaches have drawn significant attention [6] than pixel-based classification due to the availability of high resolution imagery. Pixel-based methods mainly utilize spectral information [3] and focus on the materials of individual/neighboring pixels. To correctly classify each category of weed, it is crucial to extract shape and texture information [7]. Whereas, patch-based classification method shows clear shape and texture information of the weed [4] [5]. Therefore, patch-based classification results are handy for site-specific weed management and can then be used to direct effective application of the herbicides. Hence, it reduces the amount of chemical needed to control the weeds.

It is relatively easier to use RGB images for weed classification due to limited visual spectral bands and properties. Recently, Larry et al. [4] introduced synthetic RGB images for the classification of home-lawn weeds. As the captured real images were limited, synthetic RGB images were artificially created. Hung et al. [5] used high-resolution RGB images for the classification of three different kinds of weeds. Comparing with color images, hyperspectral images consist of many contiguous and narrow bands and they provide detailed spectral characteristics within and beyond visible wavelength. This paper attempts to introduce hyperspectral images for weed classification.

Conventional feature extraction approaches require design and selection of appropriate features to achieve high classification accuracy. This laborious feature design and selection process is highly dependent on the application and type of data. Convolutional Neural Network (CNN), however, has a capacity to learn high-level spatial and spectral features, which can be expected suitable for weed detection. This is the reason we introduce CNN for weed classification and evaluate its performance. The performance of CNN is then

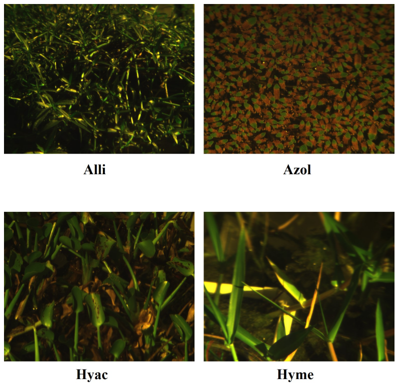


**Fig. 1**. RGB images of Alli, Azol, Hyac, and Hyme

**Table 1**. Technical Characteristics of Hyperspectral Imaging System.

| | |
|---|---|
| Model of Filter | Brimrose VA210 |
| Model of Camera | JAI BM-141GE |
| Wavelength Range | 400 - 1000nm |
| Spectral Resolution | 10 nm |
| Sensor Resolution | 1392 x 1040 pixels |
| Sensor Size | $2/3''$ |

compared to the handcrafted feature extraction method such as Histogram of Oriented Gradients (HoG). Three different patch sizes are investigated in this study. The main contribution of this paper is: CNN is introduced for spectral-spatial analysis on hyperspectral image for weed classification; we have also evaluated the performance of the CNN and HoG methods on real dataset; and show the advantages of CNN compared to HoG by using different patch sizes and the computational load.

## 2. METHODOLOGY

The methodology of this study can be decomposed into three steps: 1) hyperspectral data collection; 2) generating labelled samples, and 3) classification.

### 2.1. Hyperspectral data collection

A hyperspectral imaging system with a Brimrose VA210 filter and JAI BM-141 camera was used to gather the imagery used in this study. Its technical specifications are presented in Table 1. Each captured image has 61 bands, covering the wavelength from 400nm to 1000nm. Each band image of the hyperspectral cube has $1040 \times 1320$ pixels. One hyperspectral cube were gathered at each site, which included Hyme, Alli, Azol, and Hyac. Sample images are shown in Fig. 1. We experimented different band subsets for weed classification. Suitable patch sizes were investigated which were simulated via down-sampling.

### 2.2. Generating labelled samples

CNN models require large number of training samples to avoid over-fitting. Therefore, we increased the datasets by data augmentation methods. Two most popular and effective data augmentation methods are rotation and random cropping, which have been used in this study. Each image patch is initially rotated at five different angles and then from each rotated patch we randomly cropped number of different patches. The size of randomly cropped patches is 80-90$\%$ of the actual patch size. Using the data augmentation method, equal amount of training patches, i.e. 2000 image patches, were generated for each weed category.

### 2.3. Classification

In this study, different band subsets of the augmented labelled samples of each weed category are used to evaluate the CNN architecture and the HoG method.

#### *2.3.1. Convolutional Neural Network (CNN)*

MatConvnet tool [8] is used to train and test a CNN model. Our CNN model consists of four convolutional layers ($CNV$) and two fully connected ($FC$) layers. The size of the input images is $W_t \times H_t$ pixels. The configuration of our CNN architecture is shown in Table 2. Each convolutional layer consists of (1) Convolution layer; (2) Max-pooling layer; and (3) Nonlinearity layer [9]:

**Convolution layer**: The input to the convolutional layer is an image or a feature map $X$ of size $W_t \times H_t \times D$ with the trainable $K$ kernels, each of size $w \times h \times D$ also called the filter bank $W$. The output feature map $F_{map}$ can be computed as [9]:

$$F^K_{map} = \sum_{i=1}^{D} W_i * X^i + b \quad (1)$$

where * is a two-dimensional discrete convolution operator and $b$ is a trainable bias parameter.

**Non-linearity layer**: In the traditional CNN architecture, this layer consists of a point-wise non-linearity function, which is applied to each component in a feature map. The non-linearity layer computes the output feature map as:

$$F_{NL} = f(F_{map}) \quad (2)$$

where $f(.)$ is commonly chosen as rectified linear unit (ReLU), i.e. $f(x) = max(0, x)$.

**Pooling layer**: The pooling layer reduces the resolution of the feature map. It makes the features reliable against the noise and distortion. The pooling layer involves executing a max operation over the activations within a small spatial region $G$ of each feature map:

$$F_{PL} = \max_{i \in G} F_{NL_i} \quad (3)$$

**Table 2**. CNN Architecture for weed classification using hyperspectral dataset

| Layers | $CNV1$ | $CNV2$ | $CNV3$ | $CNV4$ | $FC1$ | $FC2$ |
|---|---|---|---|---|---|---|
| Filter size | 3x3 | 3x3 | 3x3 | 5x5 | 1x1 | 1x1 |
| Number of Filters | 64 | 128 | 256 | 512 | 1024 | 4 |

### *2.3.2. Histogram of Oriented Gradient (HoG)*

Histogram of Oriented Gradient (HoG) was introduced by [10]. As HoG features are illumination invariant and suitable to extract shape features, this algorithm is examined and compared with CNN. HoG features represent the image or each band of the hyperspectral cube by a set of local histograms. In the following we explain the sequence of steps involved in HoG feature extraction.

Initially, gradient is computed by convolving the image $I(x, y)$ with the horizontal $D_x = [-1\ 0\ 1]$ and vertical $D_y = [-1\ 0\ 1]^T$ masks which can be represented as: $G_x(x, y) = D_x * I(x, y)$ and $G_y(x, y) = D_y * I(x, y)$. where * denotes the convolution operation.

After that, the orientation $\ominus(x, y)$ and magnitude $M(x, y)$ at each pixel are computed using the ratio of gradients in vertical and horizontal directions as:

$$\ominus(x, y) = arctan\frac{G_x(x, y)}{G_y(x, y)} \tag{4}$$

$$M(x, y) = \sqrt{G_x(x, y)^2 + G_y(x, y)^2} \tag{5}$$

In the following steps, the image is divided into overlapping blocks of size $N \times N$ and then each block is divided into $M \times M$ cells. For each cell, a weighted vote is calculated for each pixel which is usually the gradient magnitude at that pixel. The votes are then accumulated into orientation bins. Finally, the HoG features are extracted from each cell and these features are equal to the number of orientation bins.

## 3. EXPERIMENTAL RESULTS

Both methods are evaluated using a multi-class hyperspectral dataset of weeds which contains one hyperspectral cube of each weed category. The dataset is first cropped from the boundaries of each hyperspectral cube to make an equal size cube of $1000 \times 1000$ pixels. As the hyperspectral images are captured very close to the targets, we down-sampled the image by averaging the $32 \times 32$ pixels to simulate lower spatial resolution. Four different datasets are created using 3, 10, 30 and 61 bands. Finally, for the investigation of three different patch sizes each dataset is divided into $500 \times 500$, $250 \times 250$, and $125 \times 125$ pixels. Extracted patches are then augmented as explained in Section. 2.2.

The datasets are randomly split for each category of weed dataset as 60% for the training set, 20% for the testing set, and 20% for the validation set. For this study, the convolutional layers of CNN architecture $CNV1$, $CNV2$, $CNV3$ and $CNV4$ are implemented with padding 1. The kernel size

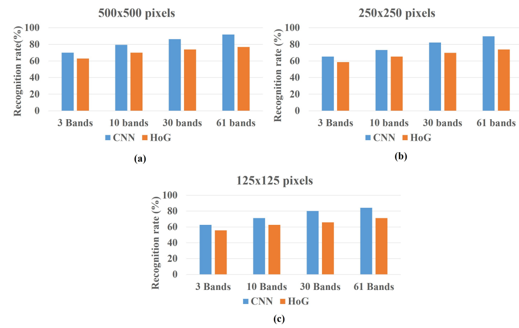


**Fig. 2**. Comparison of CNN and HoG using a patch size: a) $500 \times 500$, b) $250 \times 250$, c) $125 \times 125$ pixels for 3, 10, 30, and 61 bands)

of pooling layer is [2 ,2] with the stride of 2. CNN network is trained using the softmax classifier with the batch size of 100 and learning rate of 0.0001. Whereas, for the HoG features we used linear support vector machine (SVM) from LIBSVM [11] to train a classifier. The size input image is $56 \times 56$ for both the methods. Figs. 2-a, 2-b, and 2-c demonstrate the testing accuracy of three different patch sizes using CNN network and HoG for the dataset of 3, 10, 30, and 61 bands, respectively. It is noticed from Figs. 2-a, 2-b, and 2-c that the dataset of 3 bands (i.e., R, G, and B) does not perform very well using the CNN network and the HoG method. It is mainly because three bands could not provide significant discriminative feature information that distinguish each category of weed. However, the testing accuracy is increased by adding more bands for each category of weed using CNN. Hence, learned features using CNN give more discriminative information by adding more bands to differentiate each category of weeds. The testing accuracy for the HoG features is lower than CNN.

In this study, we also investigate three different patch sizes. It is important to use suitable size of the patch as small patch size makes it difficult to extract information. Whereas, the bigger patch size is better but it is expensive it terms of wide range of data collection. Patch size of $500 \times 500$ pixels achieved high recognition accuracy overall using CNN. Results using CNN method show that the difference between the recognition rate of 3 bands and 61 bands for the patch size of $500 \times 500$ pixels is 21.83%, 24.43% for $250 \times 250$ pixels, and 21.57% for $125 \times 125$ pixels. Overall, the patch size of $250 \times 250$ pixels achieved high recognition rate when we calculate the difference. Hence, it is suitable to extract

**Table 3**. Parameters of CNN architecture

| No. of Bands | Parameters of CNN ($M = 10^6$) |
|---|---|
| 3 Bands | 5.748M |
| 10 Bands | 5.752M |
| 30 bands | 5.763M |
| 61 bands | 5.781M |

more diverse features as compare to $500 \times 500$ and $125 \times 125$ pixels.

This study also calculates the number of parameters used by the CNN model for each dataset as:

$$P = N_k \times w \times h \times D \tag{6}$$

where $N_k$ is the number of kernels, $h$ is the kernel height, $w$ is the kernel width, and $D$ is the number of channels in the kernel.

The number of parameters directly effects the computational load of the CNN architecture [12]. Table 3 shows that the parameters used by CNN for different band subsets. The increase in computational load is not significant as the same architecture is used for all the bands. Therefore, it is found CNN is suitable for using patch-based hyperspectral images for weed classification in terms of computational load.

## 4. CONCLUSION

Weed classification using hyperspectral images is an emerging field in the area of remote sensing. In this study, we have evaluated two advanced methods for the weed classification that are originally designed for color images. Hyperspectral images are more complicated than RGB images. The evaluation of hyperspectral images using CNN network for weed classification provides a guidance to use patch-based method. Overall, results show that CNN architecture using higher number of bands achieves higher classification accuracy than HoG method. Thus, it is effective to use more detailed data information and high-level feature learning method (i.e., CNN). Also, results show that suitable patch size $250 \times 250$ pixels contains complete and adequate discriminative information. Overall, the computational load of CNN architecture is not significantly high when using more number of bands. Therefore, it is suitable for using CNN for the patch-based weed classification using hyperspectral image. For the future work, more datasets will be collected. Also, instead of using all the available bands, we intend to develop an efficient band selection and feature learning algorithm based on CNN to achieve high recognition rate with reasonable computational load for future course of action.

## 5. REFERENCES


[1] K. Smith and B. Scott, "Weed control in grain sorghum," *Grain Sorghum Production Handbook (Eds. L Espinoza, J. Kelley). Cooperative Extension Service, University of Arkansas, Little Rock, AR, USA*, pp. 47–49, 2010.

[2] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 10, pp. 6232–6251, 2016.

[3] I. Sa, Z. Chen, M. Popovic, R. Khanna, F. Liebisch, J. Nieto, and R. Siegwart, "Weednet: Dense semantic weed classification using multispectral images and mav for smart farming," *arXiv preprint arXiv:1709.03329*, 2017.

[4] L. Pearlstein, M. Kim, and W. Seto, "Convolutional neural network application to plant detection based on synthetic imagery," in *Applied Imagery Pattern Recognition Workshop (AIPR), 2016 IEEE*. IEEE, 2016, pp. 1–4.

[5] C. Hung, Z. Xu, and S. Sukkarieh, "Feature learning based approach for weed classification using high resolution aerial images from a digital camera mounted on a uav," *Remote Sensing*, vol. 6, no. 12, pp. 12 037–12 054, 2014.

[6] T. Blaschke, "Object based image analysis for remote sensing," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 65, no. 1, pp. 2–16, 2010.

[7] Y. Gao and J. F. Mas, "A comparison of the performance of pixel-based and object-based classifications over images with various spatial resolutions," *Online Journal of Earth Sciences*, vol. 2, no. 1, pp. 27–35, 2008.

[8] A. Vedaldi and K. Lenc, "Matconvnet: Convolutional neural networks for matlab." in *ACM Multimedia*. ACM, 2015, pp. 689–692.

[9] L. Zhang, L. Zhang, and B. Du, "Deep learning for remote sensing data: A technical tutorial on the state of the art," *IEEE Geoscience and Remote Sensing Magazine*, vol. 4, no. 2, pp. 22–40, 2016.

[10] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1. IEEE, 2005, pp. 886–893.

[11] C. C. Chang and C. J. Lin, "Libsvm: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, p. 27, 2011.

[12] A. Canziani, A. Paszke, and E. Culurciello, "An analysis of deep neural network models for practical applications," *arXiv preprint arXiv:1605.07678*, 2016.