# Towards Real Time Detection of Rice Weed in Uncontrolled Crop Conditions

Muhammad Umraiz[1], Sang-cheol Kim[2*]
[1]Student, Division of Electronics and Information Engineering, Jeonbuk National University
[2*]Core research institute of intelligent robots, jbnu Research Professor

# 통제되지 않는 농작물 조건에서 쌀 잡초의 실시간 검출에 관한 연구

무하마드 움라이즈[1], 김상철[2*]
[1]전북대학교 전자 · 정보공학부 학생
[2*]전북대학교 부설 지능형로봇연구소 연구교수

**Abstract** Being a dense and complex task of precisely detecting the weeds in practical crop field environment, previous approaches lack in terms of speed of processing image frames with accuracy. Although much of the attention has been given to classify the plants diseases but detecting crop weed issue remained in limelight. Previous approaches report to use fast algorithms but inference time is not even closer to real time, making them impractical solutions to be used in uncontrolled conditions. Therefore, we propose a detection model for the complex rice weed detection task. Experimental results show that inference time in our approach is reduced with a significant margin in weed detection task, making it practically deployable application in real conditions. The samples are collected at two different growth stages of rice and annotated manually.

**Key Words :** Real Time Weed Detection, Smart Farming using IoT, IoT with Object Detection, Rice Weed Detection, Weed Detection using Deep Learning

**요 약** 실제 복잡다난한 농작물 밭 환경에서 잡초를 정밀하게 검출하는 것은 이전의 접근방법들로는 이미지 프레임을 정확하게 처리하는 속도 면에서 부족했다. 식물의 질병 분류 문제가 중요시 되는 상황에서 특히 작물의 잡초 문제는 큰 화제가 되고 있다. 이전의 접근방식들은 빠른 알고리즘을 사용하지만 추론 시간이 실시간에 가깝지 않아 통제되지 않은 조건에서 비현실적인 해결책이 된다. 따라서, 복잡한 벼 잡초 검출 과제에 대한 탐지 모델을 제안한다. 실험 결과에 따르면, 우리의 접근 방식의 추론 시간은 잡초 검출 과제에서 상당한 시간절약을 보여준다. 실제 조건에서 실제로 적용할 수 있는 것으로 나타난다. 주어진 예시들은 쌀의 두 가지 성장 단계에서 수집되었고 직접 주석을 달았다.

**주제어 :** 실시간 잡초 검출, 사물인터넷을 활용한 스마트팜, 물체 검출을 이용한 사물인터넷, 쌀 잡초 검출, 딥러닝을 이용한 잡초 검출

This work was supported in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education(NRF-2019R1A6A1A09031717 and NRF-2019R1A2C1011297).
*교신저자 : 김상철(sckim7777@jbnu.ac.kr)
접수일 2020년 1월 8일 수정일 2020년 2월 13일 심사완료일 2020년 2월 17일

## 1. Introduction

Increase in hardware miniaturization capacity and power efficiency has been one of the main factors for the emergence of the Internet of Things (IoT). This domain finds tremendously successful applications in home automation systems and smart grids. Now with the rise of artificial intelligence (AI), there is need to harness the power of these domains, IoT and AI together. One such application can be realized in agriculture. In conventional agricultural framework, there has been ineffectual manpower in utilization since ages. With the rise of automation industry, vehicles like tractors oust the use of animals for the tasks like ploughing. Weeding, spreading fertilizers and watering are among the main operations in the cultivation tasks that need automation. Weeds are the potential enemies of the crops which reduce the production. To remove the weeds, spraying has been used. Whereas, even the spraying with automated commercialized vehicles is not efficient, because the process is random. It does not look for the weeds to spray on but haphazardly sprays on the plants as well [1]. On one hand, there is waste of chemicals that comes with health hazards and pollution in environment. Secondly, there is the diversity of the weeds in the crops. Therefore, we cannot spray one type of chemical for different kinds of weeds. Moreover, excessive inadequate spraying can cause the damage in the soil and affect the crop growth. Therefore, smart farming is the researchers' new area of interest now a days.

The emergence of Deep Neural Networks have revolutionized the computer vision field with its vast majority of commercial applications everywhere nowadays. The year 2012 had been a year of breakthrough for the deep neural networks when Hinton's group attained remarkable classification results on ImageNet [2,3] dataset. Since then, there has been a tremendous research in deep neural networks. Deep learning extracts features itself from the available samples without need of hand engineered features to further build classifier on top of it. Areas like recognition, object localization and 1D data is under active research of deep neural networks [4,5]. Object detection is a complex task as compared to the classification. In classification we only identify the class category given an image, whereas in object detection we have to precisely localize the position and coordinates of the object in the image. Nowadays deep learning spans almost all the technological discourses ranging from economics, finances, arts, biology and engineering. Where DNN have influenced other areas, agriculture is of no exception. Recent developments in agribots equipped with DNN which are moving into the croplands for surveillance and cultivation operations, is a step forward towards AI handling agricultural tasks [6].

Rice being one of the major source of carbohydrates, is consumed on daily basis around the globe. World's population is growing rapidly and putting pressure on the agriculture industry day by day. Now with increase in population, rice consumption is expected to increase at the same pace. Also with the expansion of cities, agriculture land is decreasing which makes it a challenge to increase production with the limited land and water resources. Rice is among one of the top food sources especially in Asia Pacific region. Like other crops, this is also prone to weeds that fight for the nutrients with the plants and cause to decrease the production by 40-60% in severe cases [7]. Furthermore, weeds also host other pests and crop diseases that are damaging for the crops. Recent advances in technology has changed the interest of the masses and they feel less inclined towards agriculture, therefore this industry is expecting less manpower in future. Various methods are being used to control the weeds that include physical methods, cultural methods and chemical methods, each having its own pros and cons. Pesticides (chemical methods) which are commonly used to remove the weeds, are creating severe environmental and health effects. Sometimes these health effects are acute that need urgent medication in case of certain chronic allergies. Therefore, some countries are bringing agricultural reforms banning the usage of pesticides. Alternatively, robotic and unmanned solutions are being investigated and widely studied in this regard. Our work is based on one type of rice weeds, called "wild millet". This weed type looks almost similar in terms of shape, color and texture with the crop and therefore sometimes hard to identify clearly even with naked eye [8]. Therefore, we propose an autonomous real time agriculture system for weeding, based on deep neural networks in the rice field which are transplanted row wise. This system is capable of identifying and localizing the rice crop and weed. We used a relatively light backbone Convolutional Neural Network (CNN) to extract the features of the input image and then based on these feature maps, we perform detection and classification of the objects. Our detection pipeline is lightweight and processes the image very fast making it practical solution for real environments. This work is part of the bigger project where we aim to build an unmanned tractor for weeding and other agriculture applications.

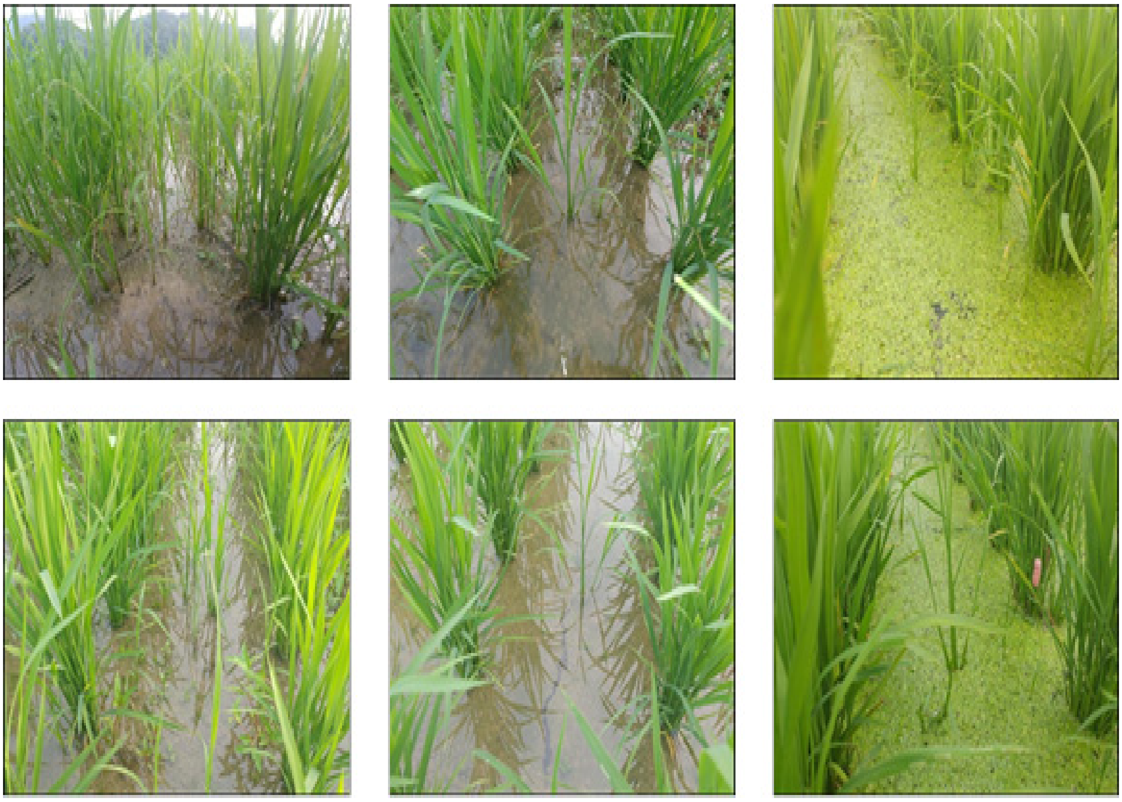
## 2. Related Work

Remote monitoring in the agriculture industry has been in practice recently by exploiting the power of Internet of Things (IoT). Sensors play a vital role in the monitoring of water-level, sunlight, moisture in the air, nutrients requirement and temperature, best suited for the specific kind of plant. Such a system was developed by JaeGu et al., where they remotely monitored the plant details through a server where data was stored and processed [9]. In the same way, Jong-Min et al. built an intelligent system to retrieve the medical images which were further used for classification purpose [10]. The proposed system was also based on Internet of Things. The application of the UAVs for plant protection has drawn attention of the many researchers recently. One study was carried in maize and sunflower with the acquisition of the weed overlays, through the method of projection calculation from spectral images at 30, 60 and 90 meters [11,12]. However, this leaves the gap for more accurate information on weed areas on a relatively small scale and makes it difficult to distinguish between crops and weeds. A recent wide spectrum study has been conducted that draws a large scale picture of the croplands risks in Australian soil and presents a robotic solution that is claimed to be applicable commercially. The research team has made the dataset of more than 17000 images publicly available calling other researchers to thrive the research domain. One such previous work was carried by Alex et al. [13]. Tallha et al. has attempted to propose a real time surveillance system of crops, exploiting parallelism in computer vision [14].

Hemming et al. utilized the machine vision based technique to localize the position of the crop plant in the rows. The detection of plant position was performed through the fast Fourier transformation [15]. Stephen et al. attempted to describe the future of the weed control [16]. Ghasemzadeh et al. developed technique to detect crop row for intra row weeding. [17]. Negrete studies the artificial vision for disease identification in crops in Mexican Land [18]. Tang et al. studied the site-specific areas spraying and developed a prototype for the robot for intra-row spraying [19]. Object detection is a multi-facet problem in which not only the recognition is performed but the pixel level location of each object inside the image is also calculated. Therefore, it is the vital step leading towards many other computer vision applications [20]. In the classic object detection pipeline 3 steps were followed which include proposal generation, feature vector extraction and region classification. This pipeline was the researchers' main area of investigation, a decade ago. Then with the passage of time, this pipeline became saturated and was replaced by modern era of object detection based on deep learning [21,22].

## 3. Dataset Collection

Millet, which is rice weed looks similar to the rice and is thus hard to distinguish. Every sample image in dataset is considered to have at least one object for each class. For real time analysis in practical environment, we collected data under two different growth stages of the rice crops under different illumination conditions that can clearly be seen in figure 1. Some samples correspond to sunny and the others correspond to cloudy conditions making it dynamic and difficult dataset to be learned by the ConvNets over texture, pattern and color. Total Images in dataset are 380, out of which, 90% samples are used for the training whereas rest of the samples are test samples. The origin of the data collection is the suburb of Jeonju-si, a regional capital of Jeollabuk-do which is a province in the Republic of South Korea. Rice has been cultivated in the form of rows, and wild millet unnecessarily grows in between the paddy lines. Domain experts have been consulted prior to the samples collection, which involves experts



[Fig. 1] Various images of the diversified Rice Millet dataset. Wild Millet can be observed ingrown between the paddy line.

from agriculture department. Rice being the major food staple consumed in the South Korea, makes this problem worth solving through deep learning, a step for transforming agriculture towards AI. Augmentation was used for only training whereas testing was performed on original images.

This dataset has many problems inherently. Since all the background including class categories are green, so this factor is enough to confuse the learning algorithm. It makes hard to focus on one particular object. Shadows of the leaves falling into the water is another factor that cannot be ignored. In some images, these shadows can also be confused with one of the categories, since weed class resembles with them the most. Weed shape is also not consistent. For example, weed leaves in some images are bent towards the ground. Occlusion is another factor that has to be dealt with. This can be observed in the figure 1 that some objects hide behind the rice leaves. We did not capture images with one particular angle that could make our approach less effective when it comes to the deployment in the field. Some of the previous approaches focus on one single angle, which is unpractical and less generalized. All these above factors make dataset of very diverse quality.

Having encountered all of the above problems, annotation comes as the major challenge, since all results are based on this. We tried different annotation methods and carried experiments, but performance remained significantly low. Since dataset is captured at two different growth stages and weather conditions, so annotating the rice class with bounding box only on leaves was not fruitful and it immediately failed. Then we tried with other positions and observed the results but they also did not work well. After experimenting for many times, we find that annotating at root positions of both categories work effectively so we adopted this approach for rest of the experiments.

## 4. Methodology

We propose one stage detector to detect rice crop weeds in the practical field environment. One stage detectors give optimal solutions when it comes to the speed and deployment. Next we describe the short explanation of 2 stage detectors and explain the proposed approach for weed detection task.

### 4.1 Two Stage Detectors

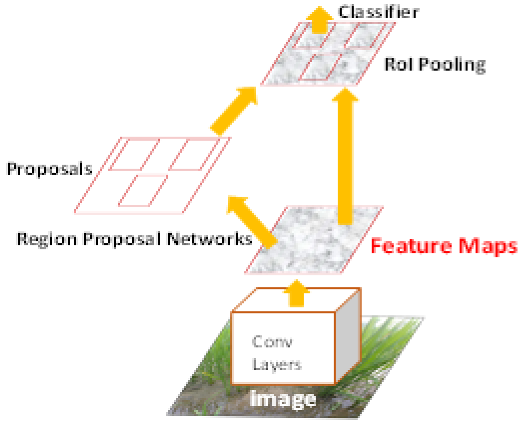
Nowadays, there are primarily two types of modern object detectors that produce state of the art results. One of them is two stage detectors and other being the single shot detectors. RCNN was among the leader boards which took off from the old detectors and gave the idea of new pipeline after resurgence of CNNs [21]. It introduced new kind of pipeline based on CNN and used another algorithm for proposal generation, called Selective Search [22]. Spatial Pyramid matching was one of the competitors in 2008, again inspired SPP-net years later, which had better training and inference speed than its counterparts [23]. Fast RCNN built on the pipeline of the RCNN and the improved structure of the SPP-net was new entry to the leader board [24]. Later the 3rd and much faster version of the two stage detectors RCNN family was Faster RCNN which shared computation across the spatial dimension with the CNN network for proposal generation [25]. It contained built in Proposal generation network called RPN (Region Proposal Network) which proposed region of interests without any third party algorithm's help.

### 4.2 YOLOv3 (Proposed Method)

The second category of the object detectors is the one which detects in one step hence got the name, single shot detectors. Since their pipeline is shorter than their counterparts, but they are relatively faster because of not having proposal generation network and are considered very sparse sliding window classifiers. Unlike RCNN family, one stage detectors view the bigger context of the image, and has ability to look for the bigger objects. Hence they deal detection as a whole regression problem. In YOLO, each image is divided into fixed number of grid cells [26]. Each of these grid cell predicts bounding boxes and then confidence score for those bounding boxes. Confidence score is defined as Pr(Object)*IoU which is zero if no object is found. IoU is the Intersection over union between the predicted and the labelled ground truth box and Pr is the probability of the presence of the object. The grid cell also predicts the C conditional class probabilities that is probability of a particular class, given the object exists in the box. There is one set of probabilities on each grid cell, no matter how many bounding boxes are predicted on this cell. Compared to the two stage detectors, YOLO's pipeline is simpler and easy to optimize and customize. One thing that comes with speed in YOLO is its low recall, which have been improved in the YOLO version 3. Fully connected layers are not present at the end stage of the YOLOv3 architecture and anchor boxes are used for prediction of bounding boxes. YOLO version 3 is improved on top of the version 2 [27,28]. It uses Darknet-53 (Network denomination) as a feature extractor CNN layers for weed and rice features, where 53 corresponds to the number of convolution layers. Darknet-53 (D-53) uses Leaky ReLu as the activation function in its convolution layers. In YOLO version 2, Darknet-19 was used but this D-53 network is improved on top of the former. This comparatively large network is more efficient than residual base networks which are composed of complex blocks leading to more computation cost. D-53 includes up-sampling, residual blocks and skip connection, still having the simple design. Now detection is at three different scales and is performed with 1x1 kernel at three different places in the architecture. Then the feature maps from 3 layers are concatenated and fused together in order to gain back the lost information from the feature maps, during the subsampling steps. Furthermore, non-maximum suppression is an important part in the detection stage for choosing the box category having maximum intersection over union and discarding those which overlap with the respective ground truth box repeatedly, with area matching more than 0.7. This threshold is changeable depending upon the application. Intersection over Union is an evaluating measure firstly used for evaluating VOC Pascal dataset, but now widely adopted in object detection problems. In the following equation for computing IoU, "A" corresponds to the area of the predicted box, whereas, "Gt" is the area denoting ground truth.

$$IoU = \frac{A \cap G_t}{A \cup G_t} \dots (1)$$

Complete network of the proposed approach is shown in the figure 2. Prediction across different scales gives YOLOv3 a boosted performance. Rice weed features are extracted based on the FPN (Feature Pyramid Network) architecture, where the last layer is responsible of predicting 3D tensor having information of class, bounding box and objectness score.Later, feature maps are merged with the previous layer and concatenated after up-sampling by factor of 2 with respect to the preceding layers. This technique allows YOLOv3 to preserve more semantic information and the following layer gets benefit from the earlier ones. Now after few more convolution layers, similar type of 3D tensor is predicted again, which is twice the size of the last predicted tensor. Consequently, for the third and final tensor, which is benefitted from the coarse grained and fine grained information flowing through the network layers, gives the final results.




[Fig. 2] The pipeline of the Two Stage detector, Faster RCNN where classification and regression is performed on top of the proposal generated by the RPN. RPN is an independent proposal generator network.

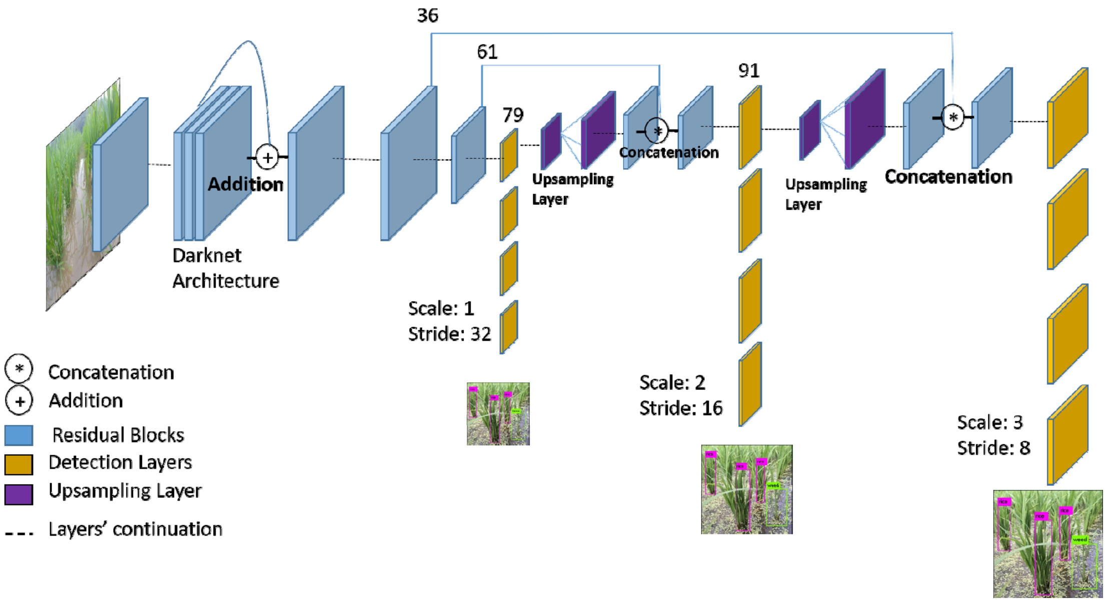
In the last two versions of the YOLO, that is v1 and v2 said to be highly struggling in predicting small objects, therefore the introduction of 3 scale prediction in 3rd version brought a significant improvement, keeping semantic information preserved. Unlike the previous two YOLO versions, YOLOv3 does not use softmax layer for class prediction. Softmax layer assumes that the object to identify are mutually exclusive and one class category cannot belong to the other, but this is not true for some datasets like in Person and Women case, which is hierarchical dataset. Therefore this is major modification in YOLOv3. Instead, predictions are made on the logistic regression layer by choosing an optimum threshold value. Predictions with certain threshold belong to that particular class. Now we explain the bounding box prediction in YOLOv3. There are 4 outputs predicted by network which represent 4 coordinates against each bounding box; *tx*, *ty*, *tw*, *th*. Now if the predicted box is offset from the top left corner of the image by (*cx*, *cy*) and bounding box prior (similar to the concept of anchor in Faster RCNN architecture) has width and height *pw*, *ph*, then the predictions are represented by following equations:

$$b_w = p_w e^{t_w} b_x = \sigma(t_x) + c_x \quad \ldots(2)$$

$$b_y = \sigma(t_y) + c_y \quad \ldots(3)$$

$$b_w = p_w e^{t_w} \quad \ldots(4)$$

$$b_h = p_h e^{t_h} \quad \ldots(5)$$




[Fig. 3] The complete architecture of YOLOv3 pipeline where detection takes place at three different scales

$$
\begin{aligned}
&Loss \\
&= \lambda_{coord}\sum_{i=0}^{S^2}\sum_{j=o}^{B} 1_{ij}^{obj}[(x_i-\hat{x}_i)^2+(y_i-\hat{y}_i)^2] \\
&+\lambda_{coord}\sum_{i=0}^{S^2}\sum_{j=o}^{B} 1_{ij}^{obj}(2-w_i\times\hat{h}_i)[(w_i-\hat{w}_i)^2]+(h_i-\hat{h}_i)^2 \\
&-\sum_{i=0}^{S^2}\sum_{j=o}^{B} 1_{ij}^{obj}[\hat{C}_i log(C_i)+(1-\hat{C}_i)\log(1-C_i)] \\
&-\lambda_{noobj}\sum_{i=0}^{S^2}\sum_{j=o}^{B} 1_{ij}^{obj}[\hat{C}_i\log(C_i)+(1-\hat{C}_i)\log(1-C_i)] \\
&-\sum_{j=o}^{B} 1_{ij}^{obj}\sum_{c\epsilon classes}^{S^2}[\hat{p}_i(c)\log(p_i(c)+(1-\hat{p}_i(c)\log(1-p_i(c)))]
\end{aligned} \tag{6}
$$

Loss function is described in multi-part in above equation (6) with 5 parts. First part represents the loss of coordinate of bounding box (central point). Second part shows the loss of height and width of the bounding box. Confidence in object with respect to the bounding box is represented in 3rd part. In the similar way 4th part shows the confidence score of having no object in the bounding box. 5th part is actually the classification loss with respect to the cell for the existence of the object. Number of cells are denoted by S2, whereas number of bounding boxes predicted by each cell are denoted by B. C represents the class categories and probability is denoted by p. λcoord and λnoobj are the hyper-parameters.

〈Table 1〉 A quick overview of characteristics of the detection models.

| Model | Faster RCNN [25] | YOLOv3 | Retina Net [29] | SSD [30] |
|---|---|---|---|---|
| Training Stages | Multiple | Single | Single | Single |
| Proposal Generation | Yes | No | No | No |
| Detector Type | Two stage | Single Stage | Single Stage | Single Stage |
| Anchor Boxes | Yes | Priors | Yes | Default Box |
| GPU Training | Yes | Yes | Yes | Yes |
| Backbone adaptability | Yes | Yes | Yes | Yes |

Table 1 shows the overview of the various characteristics of the popular object detection algorithms. Backbones are the popular pretrained feature extractors which play an important role in the detection pipeline. Without proper backbone, detection algorithms cannot extract important features of the objects. We can see in table 1 that YOLOv3 is equipped with all the vital features like other state of the art detection models. Its highly optimized Darknet-53 backbone contains skip-connection, up-sampling and residual connections. In faster RCNN, training is performed at multi-stage. Firstly, network trains the region proposal that means region proposal network learns to generate proposals given the labelled dataset. In the RPN stage, there are two heads one is for class agnostic and one for the regression. Its training takes a lot of time and this makes it less practical in environments where we have to do online training. YOLOv3 is a dense predictor and uses k-means for calculating the distribution of the priors (resembles to the concept of anchor boxes) unlike Faster RCNN where we have to train entire RPN separately. In the crop fields a robot equipped with deep learning based system cannot wait for an image to be processed and it has to move on. Therefore, Inference time reported in Table 2 show that YOLOv3 performs marginally better in such conditions proving the compatibility of the model for agriculture application.

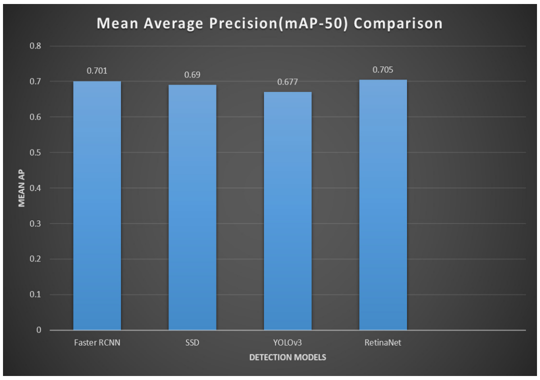
## 5. Results and Discussion

We have compared results from both type of the object detectors, one stage and two stage in figure 4. We performed the training of the network using Darknet framework. For this purpose, we used pre-trained weights of Darknet-53 network architecture and performed transfer learning for faster convergence. Darknet-53 is pre-trained on ImageNet dataset and hence captures some similar features present in Rice Crop Weed dataset. We trained model using NVIDIA Graphical Processing Unit of model Ge-Force

RTX 2080. On CPU. All the training images are in RGB domain with each having width of 804 and height of 604 pixels. We used momentum optimizer to update the weights with momentum term equal to 0.9 and decay factor of 0.0005. Since learning rate is one of the most important hyper-parameter, that's why we tweaked it often and found 0.001 as the optimum value for our weed detection task. Researchers claim that if we have a choice and time to tune only one hyper-parameter, then that should be learning rate. Similarly, we used batch gradient descent with batch size of 4 samples in each step. We trained the network for 40,000 steps. We trained all CNN layers rather than just fine tuning. It is because ImageNet does not include this weed class and hence our training could be affected. All the experiments are carried on Ubuntu 18.04 with CPU specification as Intel ® Core ™ i9-9940X CPU @3.30 GHz with 128 GB RAM.

For YOLOv3 experiments we used Darknet, a C programming language based fast library for building Deep Convolution Networks.

The above results in figure 4 show the competitive performance of the proposed method as compared to the other state of the art detection models. Mean average precision is the popular evaluating metric used to measure the performance of the detection tasks.




[Fig. 4] The mean average precision comparison among the popular two stage and one stage detectors.

〈Table 2〉 Inference time on GPU for various networks. Here FPS corresponds to the frames per second and M denotes the parameters count in millions.

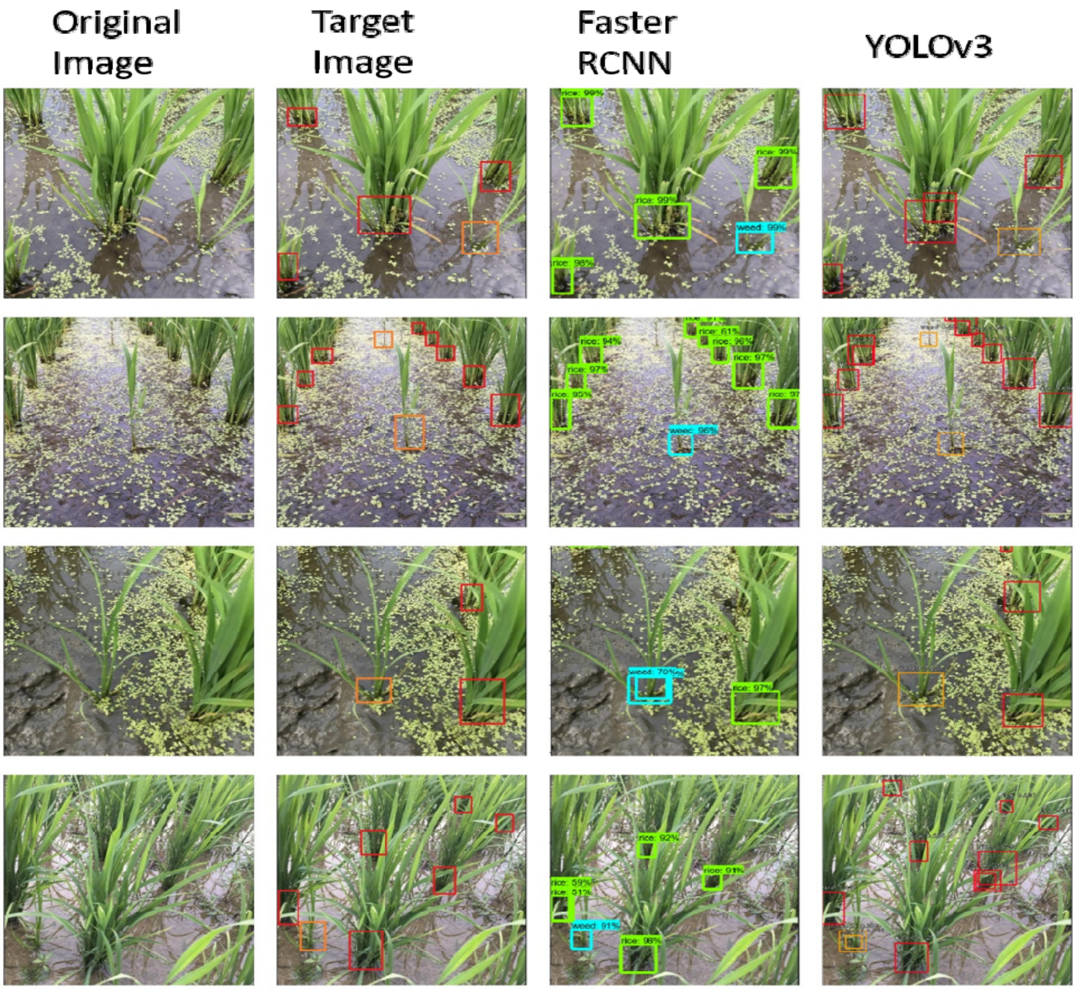
| Method | Faster RCNN [25] | YOLOv3 (Proposed) | RetinaNet [29] | SSD [30] |
|---|---|---|---|---|
| Inference time on GPU (RTX 2080) | 4.2 FPS | 40.32 FPS | 10.2 FPS | 9.54 FPS |
| # of parameters | 25.6M | 61 M | 36.4M | 25.6M |

The above table 2 describes the comparison of the different state of the art detection models. The image size is retained 804x604. YOLOv3 is ahead in speed with large margin than the other state of the art models. A deep learning system deployed in the crop field has to operate in real time though it be for the purpose of the weeding or spraying. As, this work is part of the robotic solution for weeding problem, we aim to build an unmanned system to remove the weeds from the rice crop. Now, this unmanned tractor rowing through the rice field cannot tolerate the time delay caused by the deep learning system installed over it to make some action. Only our proposed algorithm generates real time results with 40.32 frames per second with nominal precision. Hence, based on our reported result from figure 4 and table 2, we can claim that YOLOv3 is the optimum choice for our unmanned tractor application in agriculture field.

Likewise, extending this deep learning approach we can conceptualize an IoT based application also. An image acquisition system maneuvered in the agriculture field will upload the captured images to the server where proposed algorithm will process it and send back detection results in real time through internet. Based on these detection results, we can take decisions for weeding or spraying. With such an IoT system applied, multiple agriculture farms can be controlled simultaneously. Since all these actions have to be taken in real time so proposed algorithms stands best in this IoT based application as well. This approach of monitoring multiple places from a single place is economically viable also.

Figure 5 presents qualitative results of the Faster RCNN and YOLOv3. From the results analysis in figure 4, we can satisfactorily say few thing about the dataset. The dataset inherently is quite complicated because it is very difficult even with the naked eye to spot the different classes especially in those images where there is high occlusion. These quite encouraging results depict that convolution neural networks (CNNs) are one of the best architecture for learning purpose. Base models in both Faster RCNN and YOLOv3 are actually CNNs. YOLOv3 performs well in both less occluded and high occluded complex sample images.

YOLOv3 is faster in speed, and deployable but it does trade-off in the accuracy for speed, therefore in our experimental results this factor can be observed. Still YOLOv3 best fits with our application. False positives are those objects that are assigned with wrong classes after prediction, but in actual they belong to the different class. Being a dense sample detector, YOLOv3 is also prone to false detection. For the confidence




[Fig. 5] Qualitative results of proposed method for rice weed detection. (Best viewed in color) threshold (threshold chosen to assign classes to the objects)of 0.25, YOLOv3 generated 7.03% false positives which is quite promising when tested with such a complex dataset. It is because wild millet is somewhat hard to distinguish from the rice class due to its resembling contextual information. During annotation, we deliberately, kept the labelling box smaller to avoid the image covering with the whole lot of boxes. This technique was proved to be fruitful through experiments. Through experiments and careful analysis of dataset, we assume that the weed class is responsible for overall affecting the average precision.

## 6. Conclusion

In this paper we have proposed a real time deployable solution to the complex weed detection task in the uncontrolled field conditions. We proposed that YOLOv3 performs the rice weed detection task with marginal speed as compared to other detection models. We have trained and tested the models on the samples acquired manually from the practical crop field in unconditioned environments. Had it been the controlled environment, it is possible that results would have been a little different. But our main goal was to find a faster solution to the practical crop field conditions. The solution is portable and hence can be extended to other agriculture problems. Through experiments conducted on other state of the art algorithms, we prove the effectiveness of our solution. In the future work we intend to maneuver this solution to the embedded platforms like NVIDIA Jetson TX2 board, ready to apply in the field, finally making an unmanned tractor rowing between the paddy lines and detecting the weed precisely. With such a great inference time and performance, the proposed model is a competitive choice for other embedded systems and the agriculture industry.

## REFERENCES


[1] Chang, Chung-Liang, and Kuan-Ming Lin. “Smart agricultural machine with a computer vision-based weeding and variable-rate irrigation scheme,” Robotics, Vol.7, No.3, pp.38, 2018.

[2] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton, “Imagenet classification with deep convolutional neural networks,” In Advances in neural information processing systems, pp.1097-1105. 2012.

[3] Deng, J., Dong, W., Socher, R., Li, L.J., Li, K. and Fei-Fei, L., “Imagenet: A large-scale hierarchical image database,” in 2009 IEEE conference on computer vision and pattern recognition, pp.248-255, 2009.

[4] Wu, Xiongwei, Doyen Sahoo, and Steven CH Hoi. “Recent advances in deep learning for object detection.” Neurocomputing 2020.

[5] Karpathy, Andrej, George Toderici, Sanketh Shetty, Thomas Leung, Rahul Sukthankar, and Li Fei-Fei., “Large-scale video classification with convolutional neural networks,” In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, pp.1725-1732, 2014.

[6] Amer, Gulam, S. M. M. Mudassir, and M. A. Malik, “Design and operation of Wi-Fi agribot integrated system,” In 2015 International Conference on Industrial Instrumentation and Control (ICIC), pp.207-212, 2015.

[7] Hans Grinsted Jensen, Lars-Bo Jacobsen, Søren Marcus Pedersen and Elena Tavella, “Socioeconomic impact of widespread adoption of precision farming and controlled traffic systems in Denmark,” Precision Agriculture, Vol.13, No.6, pp.661-77, 2012.

[8] Mohanty, Sharada P., David P. Hughes, and Marcel Salathé, “Using deep learni ng for image-based plant disease detection,” Frontiers in plant science, Vol.7, pp.1419, 2016.

[9] J.G.Baek and H.W.Lee, “Design and implementation of self-installing Agricultural Automation System for Remote Monitoring and Control Based on LPWA Technology,” Journal of The Korea Internet of Things Society, Vol.3, No.1, pp.13-19, 2017.

[10] J.M.Kim and G.S.Ryu, “Implementation of Intelligent Medical Image Retrieval System HIPS,” Journal of The Korea Internet of Things Society, Vol.2, No.4, pp.15-20, 2016.

[11] María Pérez-Ortizac, José Peña, Pedro Antonio Gutiérrez, Jorge Torres-Sánchez, César Hervás-Martínez and FranciscaLópez-Granadosa, “Selecting patterns and features for between-and within-crop-row weed mapping using UAV-imagery,” Expert Systems with Applications, Vol.47, pp.85-94, 2016.

[12] Francisca López-Granados, Jorge Torres-Sánchez,

Ana-Isabel De Castro, Angélica Serrano-Pérez, Francisco-Javier Mesas-Carrascosa & José-Manuel Peña, "Object-based early monitoring of a grass weed in a grass crop using high resolution UAV imagery," Agronomy for sustainable development, Vol.36, pp.67, 2016.

[13] Alex Olsen, Dmitry A. Konovalov, Bronson Philippa, Peter Ridd, Jake C. Wood, Jamie Johns, Wesley Banks, Benjamin Girgenti, Owen Kenny, James Whinney, Brendan Calvert, Mostafa Rahimi Azghadi and Ronald D. White, "DeepWeeds: A multiclass weed species image dataset for deep learning." Scientific reports, Vol.9, p.2058, 2019.

[14] Tallha Akram, Syed Rameez Naqvi, Sajjad Ali Haider and Muhammad Kamran, "Towards real-time crops surveillance for disease classification: exploiting parallelism in computer vision," Computers & Electrical Engineering, Vol. 59, pp.15-26, 2017.

[15] Hemming, J., A. T. Nieuwenhuizen, and L. E. Struik, "Image analysis system to determine crop row and plant positions for an intra-row weeding machine," CIGR International Symposium on Sustainable Bioproduction, Tokyo, Japan, 2011.

[16] Young, Stephen L., and Francis J. Pierce (Eds), Automation: The future of weed control in cropping systems, Springer Science & Business Media, 2013.

[17] Hossein Behfar, Hamid Reza Ghasemzadeh, Ali Rostami, Hadi Seyedarabi and Mohammad Moghaddam, "Vision-Based Row Detection Algorithms Evaluation for Weeding Cultivator Guidance in Lentil," Modern Applied Science, Vol.8, No.5, pp. 224-232, 2014.

[18] Negrete, J. C., "Artificial Vision in Mexican Agriculture for Identification of diseases, pests and invasive plants," Journal of Advanced Plant Science, Vol.1, pp.303, 2018.

[19] Jing-Lei Tang, Xiao-Qian Chen, Rong-Hui and Miao Dong Wang, "Weed detection using image processing under different illumination for site-specific areas spraying," Computers and Electronics in Agriculture, Vol.122, pp.103-111, 2016.

[20] Yi Sun, Ding Liang, Xiaogang Wang and Xiaoou Tang, "Deepid3: Face recognition with very deep neural networks," arXiv preprint arXiv:1502.00873, 2015, Online Source https://arxiv.org/abs/1502.00873.

[21] Girshick, Ross, Jeff Donahue, Trevor Darrell, and Jitendra Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp.580-587, 2014.

[22] J. R. R. Uijlings, K. E. A. van de Sande, T. Gevers and A. W. M. Smeulders, "Selective search for object recognition," International journal of computer vision, Vol.104, pp.154-171, 2013.

[23] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," IEEE transactions on pattern analysis and machine intelligence, Vol.37, No.9, pp.1904-1916, 2015.

[24] Girshick, R., "Fast r-cnn," in Proceedings of the IEEE international conference on computer vision, pp.1440-1448, 2015.

[25] Ren, S., He, K., Girshick, R., & Sun, J., "Faster r-cnn: Towards real-time object detection with region proposal networks," in Advances in neural information processing systems, pp.91-99, 2015.

[26] Joseph Redmon, Santosh Divvala, Ross Girshick and Ali Farhadi, "You only look once: Unified, real-time object detection," in Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

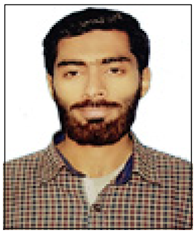
[27] Farhadi, A., and J. Redmon, "YOLO9000: better, faster, stronger," in 2017 IEEE Conference on Computer Vision and Pattern Recognition, pp.6517-6525, 2017.

[28] Redmon, Joseph, and Ali Farhadi. "Yolov3: An incremental improvement," arXiv preprint arXiv:1804.02767, 2018. Online Source https://arxiv.org/abs/1804.02767.

[29] Lin, Tsung-Yi, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár, "Focal loss for dense object detection," in Proceedings of the IEEE international conference on computer vision, pp.2980-2988, 2017.

[30] Liu, Wei, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg., "Ssd: Single shot multibox detector," in European conference on computer vision, pp.21-37, Springer, Cham, 2016.

무하마드 움라이즈(Muhammad Umraiz) [정회원]



▪ January 2017: Completed bachelor's degree in Electrical Engineering from COMSATS University Islamabad, Wah Campus, Pakistan

▪ September 2018 : Started Masters in Electronics Engineering from Jeonbuk National University.

〈Field of Interest〉

Image Processing, Deep learning for Object Detection

김 상 철(Sang-Cheol Kim) [정회원]

- February 2002 : Seoul National University, BA.
- February 2012 : Kongju National University, PhD.
- 1993 ~ December 2019 : Researcher in Rural Development Administration
- 2020 to present : Professor in Jeonbuk National University

〈Field of Interest〉

Autonomous tractor, Precision Farming, Cultivator and transplanter