Performance Evaluation of Deep Learning-based Approach for Paddy Head Detection in Images

Sujanthika Morgan[1] , Rajan S.F [1], Kaneswaran A.[1], Muralitharan R.A.[1]
[1]Department of Computer Engineering, University of Jaffna, Kilinochchi, Sri Lanka
[1]sujanthika@eng.jfn.ac.lk

Abstract: Prediction of crop yield using a computer vision-based approach is one of the active research areas in precision agriculture. Paddy head detection plays a key role in yield prediction. The performance analysis of deep learning-based models such as Faster RCNN with EfficientNet, ResNet, and YOLO to detect paddy heads available in images has been investigated in this research since rice is one of the staple foods of Asian countries. The performance of these methods is evaluated on the paddy dataset collected by the authors from the paddy lands in Sri Lanka due to the lack of a publicly available paddy dataset. Out of these deep learning-based approaches, YOLO V5 was able to achieve 88.10% accuracy.

Keywords: Paddy head detection, paddy dataset, ResNet 101, EfficientNet V2, YOLO V5, Augmentation, and precision

Introduction

Rice is the vastly cultivated crop in Asian countries which lies among the first three crops that have been cultivated by humans. They are a major source of starch energy and nutrients. The economic importance of these grain crops and their contribution to the diets of humans and livestock cannot be disputed. However, most Asian countries are involved in cultivating paddy. Due to sudden climatic changes, the farmers fail to get the expected amount of yield. This affects the food security of the world.

To compensate the food security, the production process has to be fastened. For that machines could be used for harvesting and also yield could be estimated before harvesting. For both cases paddy head detection is important.

This study uses a system that identifies and detects paddy heads from images which can improve the paddy head counting to support food security. Here we used several popular machine-learning models to detect the paddy heads. Each method used different annotation methods and underlying architecture and gave us different results. All the paddy datasets are related to Sri Lankan paddy fields.

The result of the study is the model that detects the paddy heads in an image. This can be used as a supportive tool by the government to make decisions about the quantity of paddy to ensure food security. Further research can be carried out using the dataset introduced in this research.

Related Works

Different research has been done on different datasets containing various sets of images of both wheat plants and paddy using multiple methods to identify the kernels and to predict yield. Tanuj et al. [1] used a deep-learning

network architecture for spike identification and a spikeseg Net architecture was designed which combines an encoder, decoder, and hourglass for feature extraction. The researchers have used the flood fill technique to count wheat kernels. The proposed method was able to achieve 99.91%accuracy in wheat spike detection and 95% accuracy in wheat counting however the dataset size is small which only contains 600 images. The images were taken at three different angles for one plant which is a strength. The limitation of this research [1] is that they have taken the images only at a certain growth stage of wheat so that, sometimes the crops may be destroyed.

In addition to the research [1], computer vision is also used along with deep learning in the research [4]. In this research, Tahani Alkhudaydi et al. focus on segmenting wheat spike regions by using FCN-8 architecture. The dataset contains images over the growing season from 2015 to 2017. The images contained four main growth stages of wheat which was a lacking factor in previous research [1]. The authors stated that they have used three various methods in training which involves training FCN from scratch, training FCN with two different sub-image size, and training with different growth stages. In each method, they found the accuracy reduced when using images captured during the 2017 growing season for testing rather than images captured during the 2016 growing season.

Qiongyan et al. [2] have used the color index method for plant segmentation and for spike segmentation they have used a neural network-based method using laws of texture energy. The authors focused on separating the leaves from spikes based on the energy difference. They were able to get an accuracy of 86.6%.

The algorithm developed by Qiongyan et al.[2] has been improved by Narendra Narisetti et al. [8] The algorithm that was developed by Qiongyan et al. was observed to be low in accuracy for European cultivars. Narendra Narisetti et al. introduced a method to improve the performance of the system such that it works for all wheat cultivars. Instead of original grayscale images, Narendra Narisetti et al.[8] used the wavelet amplitude as input to the laws texture energy-based neural network. Then the result of this neural network prediction has been combined with the suppressed nonâĂŞspike structures with a 5 frangi filtered image. 260 wheat cultivars were used as a dataset. This improves the performance as the algorithm can be applied to a diverse variety of wheat cultivars. They were able to produce 98.6% overall accuracy in neural network segmentation which is greater than the accuracy obtained in the research [2] done by Qiongyanet al.

Fernandez et al. [3] focused on developing an automatic ear-counting algorithm to estimate ear density. The low and high-frequency elements in the images were removed using a Laplacian frequency filter and the highest noise elements were removed with the help of a median filter. The local peaks were segmented using find maxima segmentation and it has been used to get the count of ears in the image. The top-view images of wheat during the growing season of 2014/2015 were selected for the research. The density was measured in ears/m2. The accuracy of this system was 90%.

Research carried out in [2], [4], and [8] has used a smaller number of images for training

and testing, which is a shortcoming of the research. In both [1], and [8] the researchers have considered three different side view angles when capturing images but the research [3] done by Fernandez et al. uses the images of the top view only which will cause errors when other view images were used for testing.

David et al. [5] focused on how to manipulate the wheat head dataset to get real use. The study mentioned why an extreme level of pre-processing is required to obtain wheat spikes that are visible and manipulable. In this research, RGB images have been used with different resolutions and rotations, but research [9] used images with the same resolution. Raw images were sampled up and down to get equal 1024 x 1024 squared patches. They simplified their work by using tools like Coco-Annotator [11] to label images and it will generate label information automatically. The researchers have set a baseline detection accuracy for the relevant dataset, by training a two-stage detector, Faster-RCNN, which uses ResNet34 and ResNet50 as the backbone. The input size of the images was set to 512x512 pixels to avoid memory overflow. A single image was randomly sampled into ten patches which increased the dataset to several 34220 patches. The predictions of a set of overlapping patches were merged with the results.

Hasan et al.[9] utilizes a land-based vehicle with an RGB camera to capture images of a field. In this study, they have developed a deep-learning model which detects and characterizes wheat spikes in the images. To count the wheat spikes in the images, the researchers train a variant of Region-based convolutional neural networks. But research [5] used Fast-RCNN since research [5] was conducted in mid-2019. The spikes of the images contained in the training set were labeled manually. But research [5] used an online tool called Coco-Annotator [11]. This tool annotates the wheat spikes with bounding boxes and gives the list of dimensions and coordinates of the bounding boxes during the process of training. In the end, they were able to get around 20,000 labeled spikes.

Other than research [9], [5] Changwei Tan et al. research [6] used a different method called Simple Linear Iterative Clustering. Compared to research [9], [6], this research used superpixel images to extract characteristic color parameters for initial analysis. The images were pre-processed using SLIC-based superpixel segmentation. The indices were selected using the Classification Learner in MATLAB R2016a. Both KNN and SVM were trained on the dataset. The model with the highest accuracy was chosen as the classifier. The final wheat spike was obtained using morphological processing which uses the classification results. The dilation and erosion operations were applied to the images to preserve the spikes. Here they collected 64 super-pixel images using a high-resolution camera. Researchers [9][6] used at least 3000 images for the modeling algorithm but here they used only 64 images which cover 0.75 m2 for each image.

In this paper [7], V. Crnojevic et al developed a system that detects wheat ears from thermal and RGB images. They collected images from two different wheat fields in the same geographical location. They took almost 138 RGB & Thermal images. In research [6] they used a smaller number of images like in [7] research. Their system consists of two parts a

module for the segmentation of image regions, and a module for ear counting. Here they have used three different activation functions ReLU, Parameterized ReLU & Leaky ReLu for CNN.

The research [10] by Najmah et al, modeled an algorithm for counting wheat spikes from the images. The images were acquired by a next-generation plant monitoring tool called the Crop Quant platform. Unlike research [5][9][7], here they used a special camera than a regular camera. Firstly, they transform the image data using the color index of vegetation extraction (CIVE) and then segment wheat ear regions. Then, they used Gabor filter banks and K- means clustering algorithm to detect wheat spikes. Finally, they used the regression method to estimate the number of wheat spikes.

A deep learning model called Deep Count was used to count wheat ears investigated by Sadeghi et al [12] using CNN. 121 images were acquired from 3 previous experiments [13] carried out at Rothamsted Research, UK during 2014-2016. The objective of this research was to a model system that automatically quantifies the number of spikes in the images and then calculates the number of ears per square meter. They have used the SLIC model to get the best results.

Tan et al. [14] used simple linear iterative clustering on superpixel segmentation of the images of field-grown wheat to count wheat spikes. They have extracted characteristic color parameters and used some classifiers SVM and KNN for image classification. The morphological transformation was used to extract wheat spikes the inflection points of the backbone architecture were used to determine the number of wheat spikes. The proposed method has experimented on images captured under four different concentrations of nitrogen fertilizer added to wheat plants. The accuracy was above 90% in all four conditions however only 64 images were used. The authors mentioned that this method cannot be used during poor growth status and when heterogeneity is limited.

Korohou et al. estimated the yield of wheat using the regression models in this research [15]. The regression models are linear SVR, quadratic SVR, cubic SVR, RBF SVR, and linear regression. The author was able to obtain an R2 of 0.9893 and RMSE of 0.0684 mm which was seen to outperform all the models. The research was carried out on two different wheat cultivars hence this method can be used on other wheat cultivars too. The dataset contained 1500 samples, where images were taken at every 120° orientation for each wheat plant however images 8 were taken after harvesting and the wheat plants were aligned for imaging purposes. So, the performance of the proposed method can be low when the plant background was added.

Cointault and Bernard [16] proposed a color hybrid space for yield prediction by counting wheat ears in a semi-automatic method. They were able to conclude that a rough yellow-green object is a wheat head, a green object is considered a stem or a leaf, and an object without green or yellow is considered soil based on the color study. Discriminant analysis and texture analysis were used while training and mathematical morphology were used for ear counting. The accuracies ranged from 73% to 85% in head recognition.

In this research [17] Xiong et al. use Tasselnet, a convolutional neural network-based local

regression model to count wheat spikes. The method incorporated CNN to increase efficiency. The proposed method was able to acquire an accuracy of 91.01% with a large-scale wheat spike-counting dataset that contains 1,764 images. The images were captured at three different times that is: morning, afternoon, and evening on the same day to avoid the effects of illumination levels. The limitation of this method is that there are errors when detecting overlapping wheat kernels.

Objectives

The objective of this research is to detect the paddy heads from the images that were captured directly from the field, where this would help design harvesting machines that could detect the paddy heads correctly in a real-time application and also further it could be used to predict the yield of a paddy field before harvesting. Our model could be used to detect the paddy heads of all growing seasons and of all angles from a set of images that are taken from the paddy field itself. There are different Geno types of rice available around the world in different sizes. With relevant information about the genotypes and the climatic conditions, our model can be improved to provide a solution that can help to pre-estimate the yield before harvesting.

Methodology

The study has been carried out in four phases to evaluate the performance of different deep-learning models for paddy head detection. We have identified phases. Phase I deals with dataset collection for the research. Phase II deals with data pre-processing for training. Phase III mainly deals with training models using the pre-processed data from the collected datasets and testing the trained models. Phase IV focuses on finding the best possible model for paddy head detection using the preprocessed images using some selected performance metrics.

Phase I: Dataset Collection

For this research, we have collected paddy images from the paddy fields of different parts of Sri Lanka. The paddy images have been captured between 1m to 3m above the ground level from paddy fields located in Kurunegala, Kilinochchi, Mullaitivu, and Matalw3sze districts. All the images were collected at the last growing stage of the paddy near harvesting from different genotypes. The images were captured in both morning and evening time to avoid the effects of light. Figure 1 shows the sample images from the dataset.

Figure 1: I.Sample images of the dataset

Data Preprocessing

To interpret the images, a manual inspection was done. From this inspection blurred images, high contrast images, images having an overlap of plants, and images where no grain heads are visible were identified and removed. The resolution of the images varied according to the places where the images were captured. Therefore, images were rescaled to maintain a more similar resolution. The images were

cropped into 412 × 412 pixels square patches. After that, we did some series of augmentation to the images. We started with rotating images by 90°, blurred, flipped horizontally & vertically, zoomed in at different levels, sheared 15° vertically and horizontally, and adjusted the brightness between -25% to 25%. Exposure was also adjusted between - 25% and 25%. In the end, we got a 4 times larger dataset than before. After augmentation, there were around 816 images in the dataset.

Then the images were annotated in such a way that the annotation covers only the paddy by connecting all the outer points of the paddy heads as shown in Figure 2.

Figure 2: Phase II: Building and training paddy head detection model

Deep learning-based methods have been built to detect paddy heads. While searching for a solution we were able to find an approach known as a Residual Network (ResNet) where a neural network is used as the backbone. When we add more layers to the neural network, it becomes more difficult in training, and also the accuracy starts to saturate at a point and then decreases gradually. When considering these problems ResNet becomes the best choice. The performance of classification tasks could be improved using ResNet. There are shortcut connections in the ResNet that allow the information across layers without attenuation, which improves optimization as well as reduces the difficulty in training.

We have trained the dataset with several forms of ResNet. As the baseline ResNet 50 was trained on the dataset. With the help of baseline implementation, the dataset was trained on ResNet 101 with cross-validation. Then a new technique Efficient net version 2 is used to train the model. The different architecture of efficient net v2 was used for this purpose.

As the accuracy from both ResNet and Efficient Net was not satisfied so, we moved to a better model which is YOLO v5. YOLOv5 outperformed both the above-mentioned models with the highest mean average precision of 92.68% at a threshold of 0.45. Figure 3 shows sample detected paddy heads using the YOLO v5 model with confident scores.

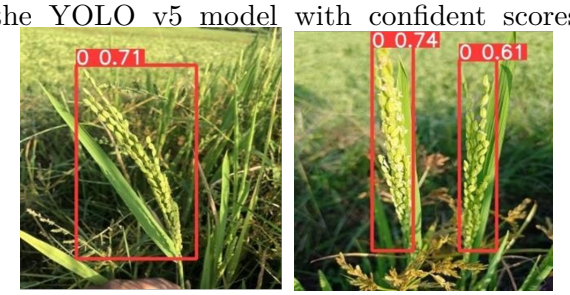

Figure 3: Phase II: Building and training paddy head detection model

Results & Discussion

For paddy head detection, several models have been investigated such as Faster RCNN with ResNet50 backbone, Faster RCNN with ResNet101 backbone, Efficient net V2 with B0 architecture, Efficient. net V2 with B1 architecture, Efficient net V2 with B2 architecture, Efficient net V2 with B3 architecture and YOLOV5. Table 1 shows the performance of different deep learning-based approaches for paddy head detection.

Table 1: Performance of paddy head detection models

Model	Precision (%)
Faster RCNN with ResNet50 backbone	82.76 +/- 2.54
Faster RCNN with ResNet101 backbone	85.96 +/- 1.76
EfficientNet V2 with B0 architecture	79.93
EfficientNet V2 with B1 architecture	79.97
EfficientNet V2 with B2 architecture	80.03
EfficientNet V2 with B3 architecture	80.39
YOLO V5	mAP@0.4592.68

Faster RCNN with ResNet50 backbone is the baseline model of our system. We were able to achieve a mean accuracy of 82.76% with s standard deviation of 2.54%, when increasing the number of epochs and also while implementing the cross-validation method. An improved backbone of ResNet101 gives a mean accuracy of 85.96% with a standard deviation of 1.76%. We used four versions of EfficientNetV2 to improve the precision. But unexpectedly we got less precision from 79.93% to 80.39% even though EfficientNetV2 was a state-of-the-art backbone.

Both ResNet and EfficientNetV2 use bounding boxes as an annotation. These bounding boxes cover a lot of extra spaces along with paddy heads which are not the area of interest for the research. Due to this reason, we had to find a better annotation method that connects all the outer points of the paddy head as in Figure 2 which can be used to train YOLOV5. YOLOV5 can be used to train and test the dataset with custom annotations. The paddy heads are annotated along with their shape which ensures that no extra spaces are covered in the annotation we used before. YOLOV5 gives 88.10% of mean average precision which was the best of the other models we used.

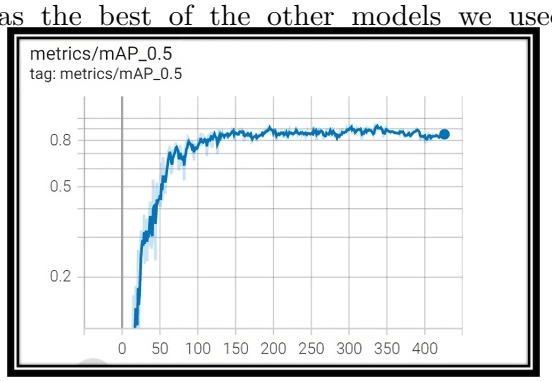

Figure 4: Mean Average Precision Vs Epochs of YOLOv5

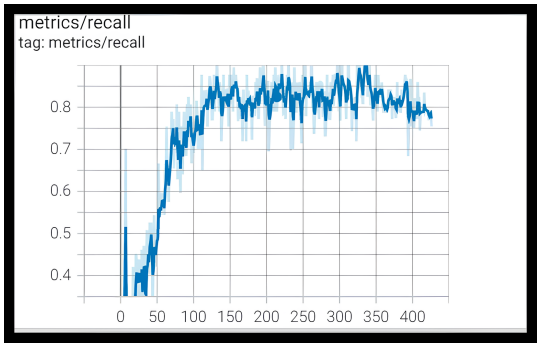

Figure 5: Recall Vs Epochs of YOLOv5

While collecting the dataset, we faced issues like continuous wind blowing & overlapped images. So that suitable image for training was very low in number. Since paddy images were very clumsy, it was very hard to preprocess the images for the training. Due to that, we had to remove unsuitable images from the dataset. The YOLOV5 used annotation which is the exact shape of paddy heads. This made model performance better and also made annotation a complex task.

Conclusions

Paddy head detection is an important method in finding paddy production and crop management. So we experimented with finding the most appropriate model for paddy head detection with the highest accuracy. From the results, we may conclude that YOLO v5 outperformed all the other models with a mean average precision of 88.10%. The highest precision was obtained because of the annotation of images where other models accept only a bounding box and YOLO v5 does not restrict to only a bounding box. Also YOLO v5 uses an auto anchoring method, where anchors are the boxes with a different aspect ratio of the object classes. First, the YOLOv5 model will fit the dataset to a particular aspect ratio box which is an anchor, and if there was no improvement in performance then it automatically changes the anchor. In this way, the YOLOv5 can produce the best results.

References

[1] Misra, Tanuj, Alka Arora, Sudeep Marwaha, Viswanathan Chinnusamy, Atmakuri Ramakrishna Rao, Rajni Jain, Rabi Narayan Sahoo, et al. "SpikeSegNet-a deep learning approach utilizing an encoder-decoder network with an hourglass for spike segmentation and counting in the wheat plant from visual imaging." Plant Methods 16, no. 1 (2020): 1-20.

[2] Qiongyan, Li, Jinhai Cai, Bettina Berger, Mamoru Okamoto, and Stanley J. Miklavcic. "Detecting spikes of wheat plants using neural networks with Laws texture energy." Plant Methods 13, no. 1 (2017): 83.

[3] Fernandez-Gallego, Jose A., Shawn C. Kefauver, Nieves Aparicio GutiÃĺrrez, MarÃŋa Teresa Nieto Taladriz, and JosÃĺ Luis Araus. "Wheat ear counting in-field conditions: high throughput and low-cost approach using RGB images." Plant Methods 14, no. 1 (2018): 22.

[4] Alkhudaydi, Tahani, Daniel Reynolds, Simon Griffiths, Ji Zhou, and Beatriz De La Iglesia. "An exploration of deep-learning based phenotypic analysis to detect spike regions in field conditions for UK bread wheat." Plant Phenomics 2019 (2019): 7368761.

[5] David, Etienne, Simon Madec, Pouria Sadeghi- Tehran, Helge Aasen, Bangyou Zheng, Shouyang Liu, Norbert Kirchgessner, et al. "Global Wheat Head Detection (GWHD) dataset: a large and diverse dataset of high resolution RGB labeled images to develop and benchmark wheat head detection methods." arXiv preprint arXiv:2005.02162 (2020).

[6] Tan, Changwei, Pengpeng Zhang, Yongjiang Zhang, Xinxing Zhou, Zhixiang Wang, Ying Du, Wei Mao, Wenxi Li, Dunliang Wang, and Wenshan Guo. "Rapid Recognition of Field-Grown Wheat Spikes Based on a Superpixel Segmentation Algorithm Using Digital Images." Frontiers in Plant Science 11 (2020): 259.

[7] Grbovic, Zeljana, Marko Panic, Oskar Marko, Sanja Brdar, and Vladimir Crnojevic. "Wheat Ear Detection in

RGB and Thermal Images Using Deep Neural Networks." environments 11, no. 12 (2019): 13.

[8] Narisetti, Narendra, Kerstin Neumann, Marion S. RÃűder, and Evgeny Gladilin. "Automated spike detection in diverse European wheat plants using textural features and the Frangi filter in 2D greenhouse images."Frontiers in Plant Science 11 (2020): 666.

[9] Hasan, Md Mehedi, Joshua P. Chopin, Hamid Laga, and Stanley J. Miklavcic. "Detection and analysis of wheat spikes using convolutional neural networks." Plant Methods 14, no. 1 (2018): 100.

[10] Alharbi, Najmah, Ji Zhou and Wenija Wang. "Automatic Counting of Wheat Spikes from Wheat Growth Images."ICPRAM (2018)

[11] https://github.com/jsbroks/coco-annotator

[12] Sadeghi-Tehran, Pouria & Virlet, Nicolas & Ampe, Eva & Reyns, Piet & Hawkesford, Malcolm. (2019). DeepCount: In-Field Automatic Quantification of Wheat Spikes Using Simple Linear Iterative Clustering and Deep Convolutional Neural Networks. Frontiers in Plant Science (2019)

[13] Virlet, N., Sabermanesh, K., Sadeghi-Tehran, P., and Hawkesford, M. J. Field Scanalyzer: an automated robotic field phenotyping platform for detailed

[14] Tan, Changwei, Pengpeng Zhang, Yongjiang Zhang, Xinxing Zhou, Zhixiang Wang, Ying Du, Wei Mao, Wenxi Li, Dunliang Wang, and Wenshan Guo. "Rapid Recognition of Field-Grown Wheat Spikes Based on a Superpixel Segmentation Algorithm Using Digital Images." Frontiers in Plant Science 11 (2020): 259.

[15] Korohou, Tchalla, Cedric Okinda, Haikang Li, Yifei Cao, Innocent Nyalala, Lianfei Huo, MouloumdÃĺma Potcho, Xiang Li, and Qishuo Ding. "Wheat Grain Yield Estimation Based on Image Morphological Properties and Wheat Biomass." Journal of Sensors 2020 (2020).

[16] Cointault, FrÃĭdÃĭric, and Bernard Chopinet. "Colour- texture image analysis for in-field wheat head counting." In Proceedings. 2nd. Symposium on Communications, Control and Signal Processing (ISCCSP)(Marrakech). 2006.

[17] Xiong, Haipeng, Zhiguo Cao, Hao Lu, Simon Madec, Liang Liu, and Chunhua Shen. "TasselNetv2: in-field counting of wheat spikes with context-augmented local regression networks." Plant Methods 15, no. 1 (2019): 150