Wheat Head Detection using Convolutional Neural Networks for Yield Monitoring

B.Tech Project Part - 1

submitted by

Aviral Jain

(18AG3AI08)

under the guidance of

Dr. Rajendra Machavaram

Agricultural and Food Engineering Department

Indian Institute of Technology Kharagpur

Kharagpur, West Bengal, India - 721302.

September - November 2021

Wheat Head Detection using Convolutional Neural Networks for Yield Monitoring

B.Tech Project Part - 1

submitted by

Aviral Jain

(18AG3AI08)

approved by

M. Rajendra

Dr. Rajendra Machavaram
(**Project Supervisor**)
25-11-2021

Prof. Rintu Banerjee
(**Head of the Department**)

Agricultural and Food Engineering Department

Indian Institute of Technology Kharagpur

Kharagpur, West Bengal, India - 721302.

September - November 2021

Table of Contents

List of Figures

Declaration

I certify that

(a) The work contained in this thesis is original and has been done by me under the guidance of my supervisor Dr. Rajendra Machavaram.

(b) The work has not been submitted to any other Institute for any degree or diploma.

(c) I have followed the guidelines provided by the institute in preparing the thesis.

(d) I have confirmed the norms and conditions given in the Ethical Code of Conduct of the Institute.

(e) Whenever I have used materials (data, theoretical analysis, figures, and text) from other sources, I have given due credit to them citing them in the text of the thesis and giving their details in the References section.

Date: 24/11/2021

Place: Kharagpur

Aviral Jain

(Student)

1. Introduction

1.1 General

With the rising global population, and ever increasing food security concerns rising, the continued development of global wheat is crucial for long-term food security. Wheat crop is India's prime most staple harvest, placed second only to rice. It is mostly consumed in the north and north-west parts of the country. Being rich in protein, vitamins and carbohydrates, it provides a balanced food to millions of people each day. New plant-breeding techniques enable the creation of new wheat plant types with desired characteristics such as disease resistance, climate resistance, and higher yields. Wheat breeding, on the other hand, is mostly done in the old-fashioned approach, which is virtually entirely manual and thus prone to error. Wheat head numbers per unit ground area, among all attributes, is an important yield component that is still manually examined in breeding trials. Wheat head count is a new decision-making tool used by plant breeders to select which wheat varieties should be crossed to produce a new, superior progeny.

It is also the main component of estimating the crop yield potential. Yield Estimation can be useful in the following ways:

- Determine fertilizer recommendations
- Estimate nutrient removal
- Making replanting decisions
- Crop Insurance purposes
- Developing pest management recommendations
- Harvest and inventory planning
- Sale and Marketing decisions

Hence, an automated wheat head detection method is required to assist the wheat head detection task that can help estimate the yield of the crop.

1.2 Justification

The number of wheat heads per unit area is an important component of estimating the crop yield potential. Not only just yield, it also helps in determining pertinent wheat traits such as head population density and head characteristics like health, size, maturity stage, presence of awns, etc.

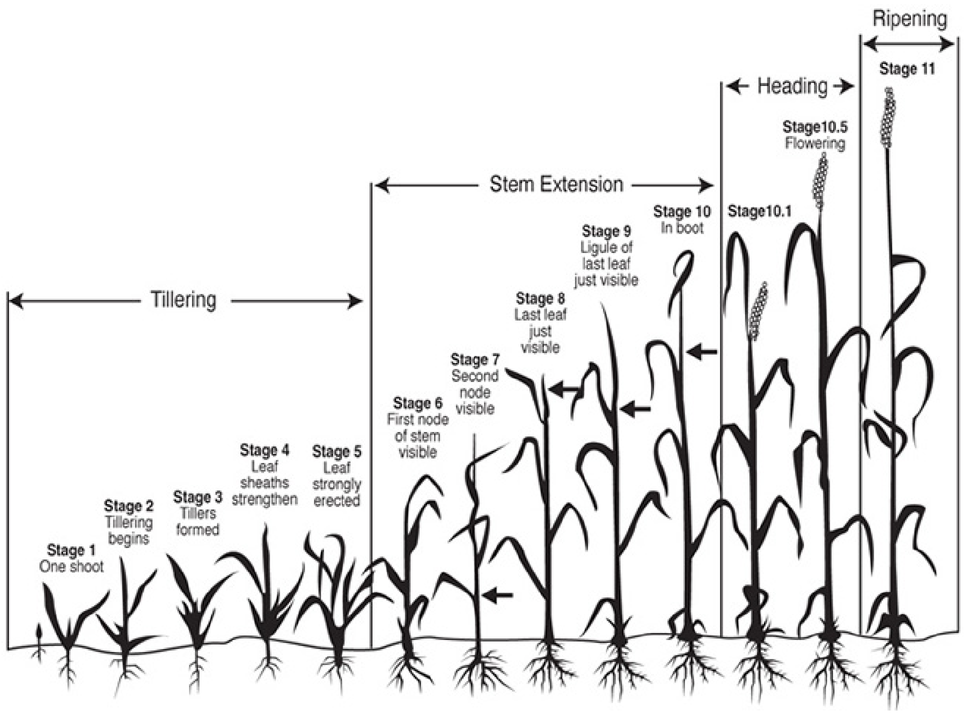


Fig 1.1 Growing stages of wheat, (courtesy: Google Images)

After 12-22 weeks of sowing, i.e. when the wheat is **headed**, it is possible to make a good estimate of the crop's yield using the count method. The study shown in **[4]** tells us some of the ways to estimate the wheat crop yield using headcount. Some of them are:

1. Based on seed weight

$$Yield\ (ton/ha) = \frac{Heads\ (per\ m^2) \times N_{grains\ per\ head} \times W_{seed\ (est.)}(mg)}{100{,}000}$$

(Approximate wheat seed weighs 30-45mg)

2. Based on row spacing

$$Yield\ (bushels/acre) = \frac{Heads\ (per\ ft.) \times N_{kernels\ per\ head}}{Row\ Space\ (inches)} \times\ 0.48$$

(1 wheat bushel = 36.74 ton)

The number of kernels per head depends upon the variety of seed used. The GWHD uses more than 8 different types of seed variety from all around the world, making it easier to generalize the study. Normally, there are between 18-26 kernels per head.

Row spacing is determined through a variety of factors. High spacing allows for better moisture absorption and easier stubble management. On the other hand, it causes lower yield potential under high yield conditions, and reduces crop establishment when applying nitrogen fertiliser with the seed. On average, it is suggested to keep about 22.5cm of row spacing.

Accuracy of yield estimates depends upon an adequate number of counts being taken so as to get a representative average of the paddock. The yield estimate determined will only be a guide and assumptions made from the estimates will contain a degree of uncertainty. Hence to aid working with a large sample, computerized methods for detection are necessary to reduce man hours involved.

1.3 Objectives

1. To identify the important attributes of the Wheat Head Detection from the GWHD Dataset

2. To develop a Faster-RCNN and EfficientDet model for the detection of Wheat heads and yield estimation.

3. To compare the performance of developed models for yield estimation in terms of accuracy and computational time.

2. Literature Review

David et al. (2020) indicated that the issue of detecting wheat heads in plant photos is critical for assessing crucial wheat attributes including head population density and head characteristics like health, size, maturity stage, and awn presence. Several studies had used machine learning techniques to build approaches for detecting wheat heads from high-resolution RGB photography. These approaches, on the other hand, have often been calibrated and verified on small datasets. Wheat head detection is difficult for computer vision because of the wide range of observing settings, genotypic variances, developmental phases, and head orientation. This task is made significantly more difficult by the possibility of blurring due to motion or wind, as well as overlap between heads in dense crowds. The Global Wheat Head Detection (GWHD) dataset is a big, diversified, and well-labeled dataset of wheat pictures that was created through a joint international collaborative effort. It includes 4700 high-resolution RGB photos and 190000 labelled wheat heads from various countries throughout the world at various stages of growth and genotypes.

***Lyon* et al. (2007)** in their research talk about several methods to calculate the wheat yield potential. They use several crop attributes like row spacing, average number of heads per foot, average number of kernels per head, etc. They also describe in detail the steps to follow and the precautions to take while conducting the experiment.

Zhao et al. (2014) presented a comprehensive overview of deep learning-based object recognition frameworks that handle various sub-problems, such as occlusion, clutter, and poor resolution, with varying degrees of R-CNN modification. The discussion begins with generic object detection pipelines, which serve as the foundation for other related activities. The breakthroughs in neural networks and related learning systems, which provide essential insights and guidelines for future progress, have been the focus of this review.

Girshick et al. (2015) proposed Faster R-CNN, presenting RPNs for efficient and accurate region proposal generation. Their approach shares convolutional features with the down-stream detection network, hence making the region proposal step nearly cost-free. Their method enables a unified, deep-learning-based object detection system to run at near real-time frame rates. The learned RPN also improves region proposal quality and thus the overall object detection accuracy.

Redmon et al. (2016) proposed YOLO, a new approach to object detection. Classifiers have been repurposed to do detection in previous work on object detection. Rather, we consider object detection to be a regression issue with spatially separated bounding boxes and associated class probabilities. In a single assessment, a single neural network predicts bounding boxes and class probabilities directly from entire images. Because the entire detection pipeline is a single network, it can be optimised directly on detection performance from beginning to end. This makes the computation extremely fast as compared to others.

Minxing et al. (2020) proposed a weighted bi-directional feature pyramid network (BiFPN), which allows easy and fast multiscale feature fusion. They also developed a compound scaling method that uniformly scales the resolution, depth, and width for all backbone, feature network, and box/class prediction networks at the same time. Based on these optimizations and better backbones, they finally made a new family of object detectors, called EfficientDet, which consistently achieve much better efficiency than prior art across a wide spectrum of resource constraints.

3. Materials and Methods

3.1 Dataset

The standard for wheat head detection is the GWHD (Global Wheat Head Detection) dataset [1]. From 2016 to 2019, data was collected from nine research organisations in various places, totaling 507 genotypes from Europe, North America, Australia, and Asia. RGB photos from a number of field-based phenotyping systems with cameras are included in the GWHD collection. The GWHD dataset was gathered through the Kaggle platform's global wheat detection competition. Because the test dataset's ground truth isn't publicly available, we divided the 3422-image training dataset into 80 percent and 20 percent training and validation sets, respectively.

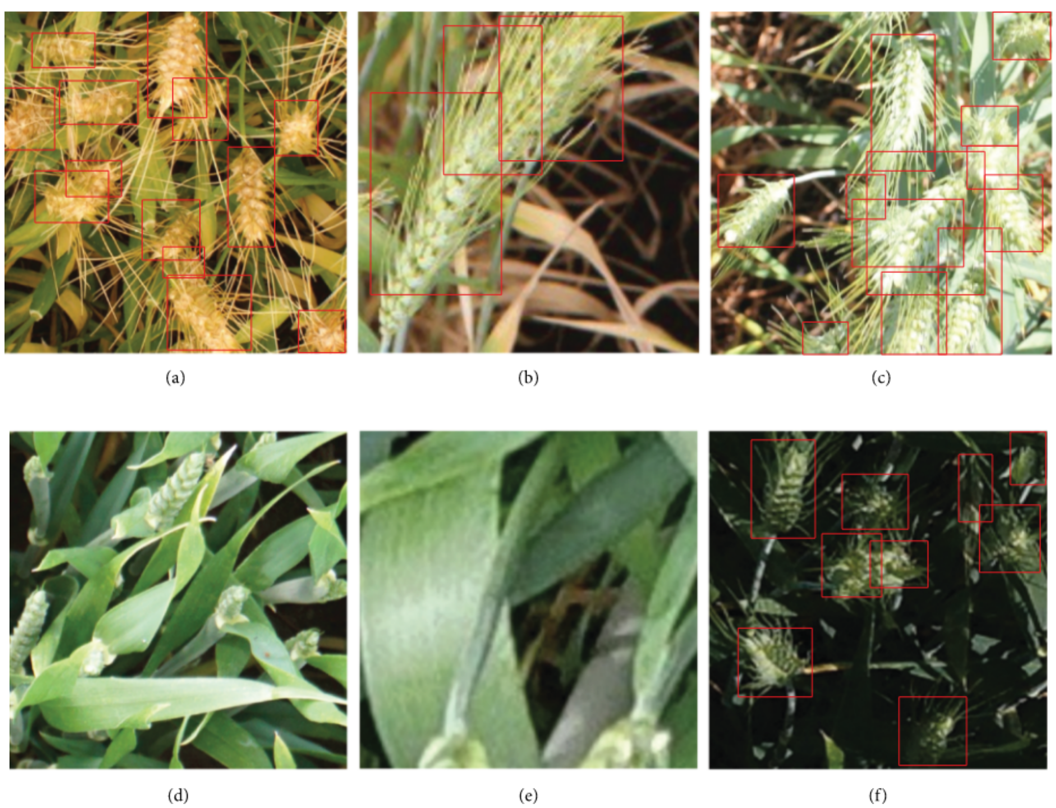


Fig 3.1.1 An example of images from the GWHD dataset

	image_id	width	height	bbox	source	boxes	box_count	per_area	max_area	brightness
0	b6ab77fd7	1024	1024	[834.0, 222.0, 56.0, 36.0]	usask_1	[[834.0, 222.0, 56.0, 36.0], [226.0, 548.0, 13...	47	36.395550	2.788544	28.905823
1	b53afdf5c	1024	1024	[988.0, 781.0, 36.0, 96.0]	usask_1	[[988.0, 781.0, 36.0, 96.0], [331.0, 863.0, 70...	46	32.459259	2.102852	32.344990
2	7b72ea0fb	1024	1024	[332.0, 662.0, 113.0, 50.0]	usask_1	[[332.0, 662.0, 113.0, 50.0], [285.0, 755.0, 3...	41	29.572105	2.300262	26.889556
3	91c9d9c38	1024	1024	[124.0, 273.0, 59.0, 73.0]	usask_1	[[124.0, 273.0, 59.0, 73.0], [688.0, 939.0, 61...	33	20.505047	2.112770	28.044238
4	41c0123cc	1024	1024	[0.0, 669.0, 73.0, 111.0]	usask_1	[[0.0, 669.0, 73.0, 111.0], [572.0, 757.0, 110...	34	38.740921	13.862610	27.353562

Fig 3.1.2 A brief overview of what the training data frame looks like. We can see that the 'boxes' column has a list of bounding boxes present in an image in the format (x, y, w, h)

Following are some of the observations from our exploratory data analysis:

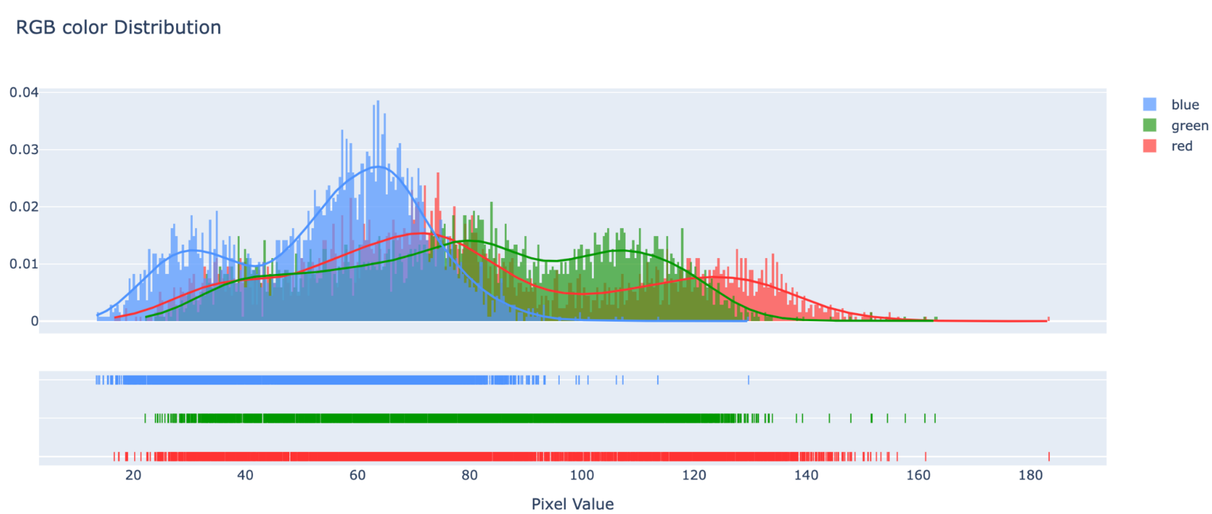


Fig 3.1.3 shows an average RGB histogram of the images from the training dataset, on the x-axis are the brightness values (0-255) and on the y-axis are the probability of the number of pixels in that region.

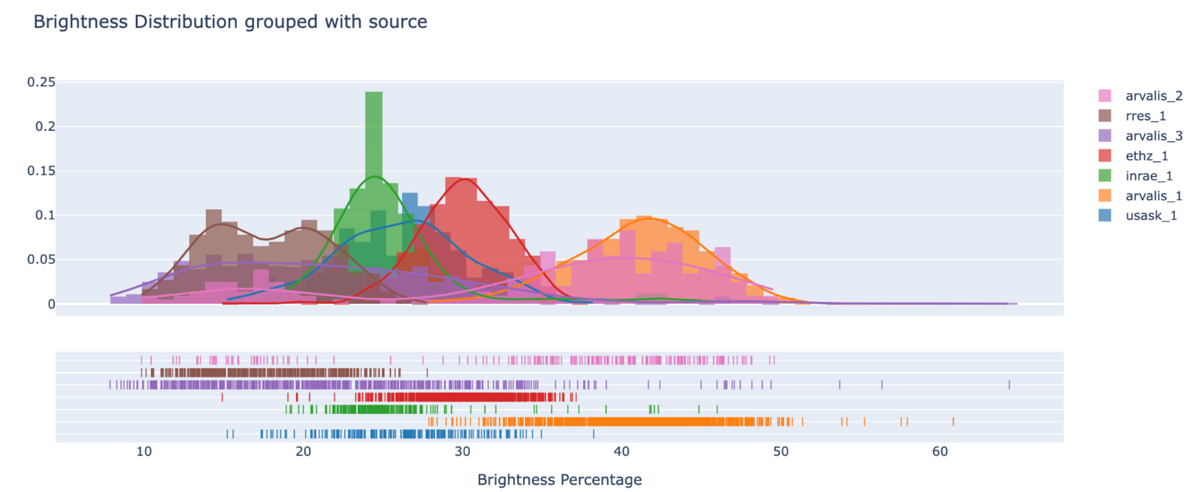


Fig 3.1.4 shows a source wise probability distribution of the pixel brightness in an image; we can see that sources like aravlis_1 mostly contribute lighter images, while rres_1 contributes only darker images.

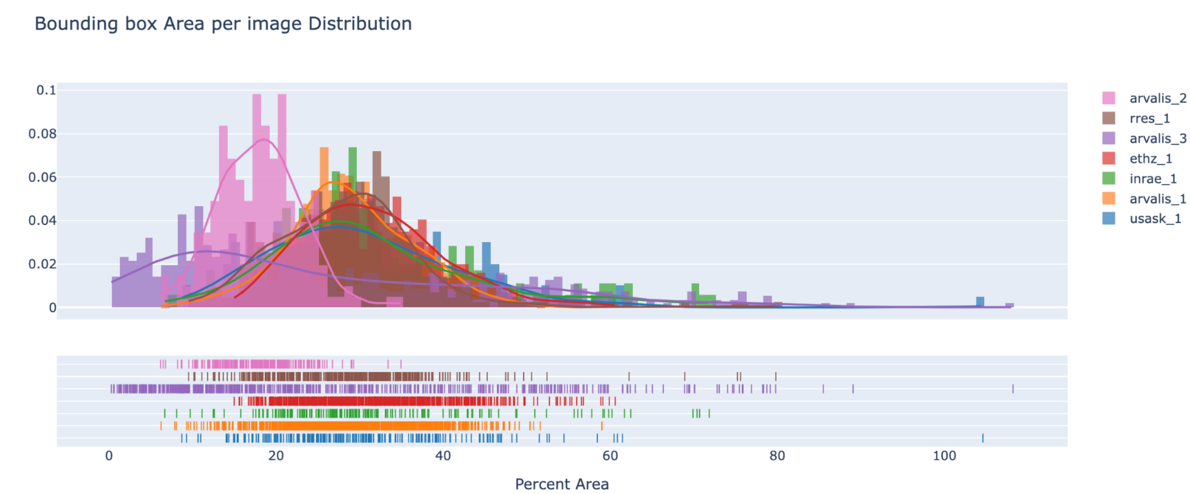


Fig 3.1.5 shows a source-wise probability distribution of the area of bounding boxes in an image. The distribution follows a slightly skewed Gaussian curve with median ranging between 20-35% area depending upon the source.

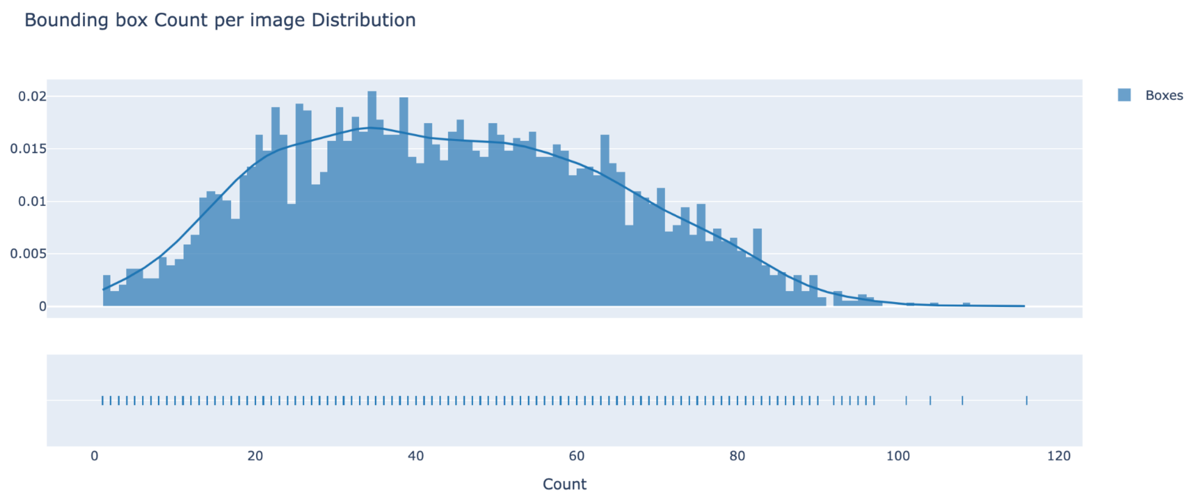


Fig 3.1.6 shows a probability distribution of the number of bounding boxes in an image, average number of bounding boxes in an image is found to be 43.8

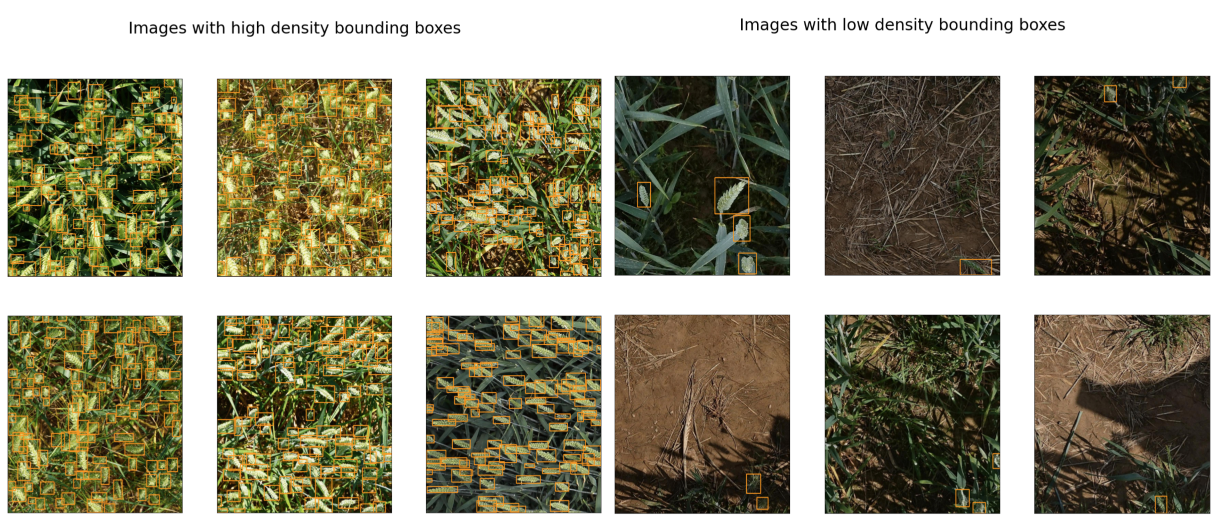


Fig 3.1.7 Images with a high density of bounding boxes (left) vs images with a low density of bounding boxes (right)

Fig 3.1.8 Images containing high area bounding boxes (left) vs images with low area bounding boxes (right)

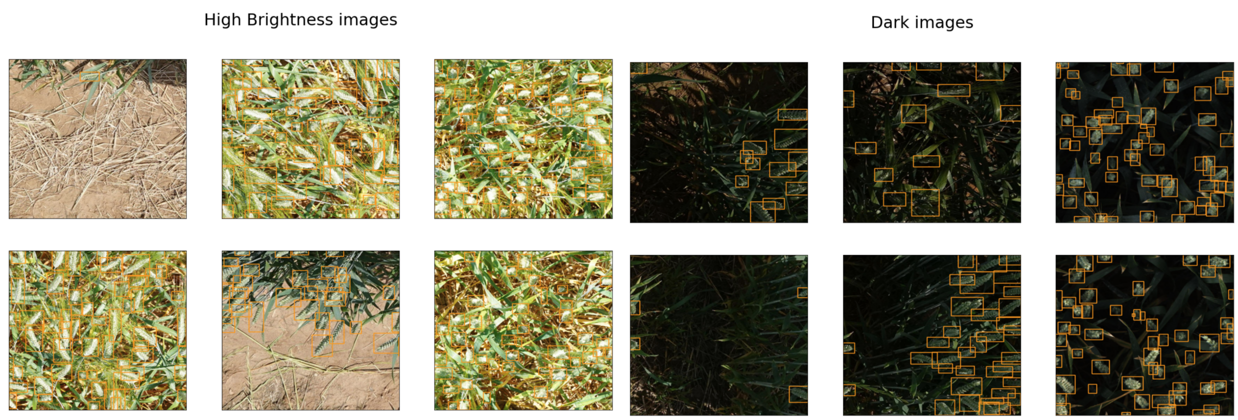


Fig 3.1.9 Example of high brightness images (left) vs low brightness images (right)

3.2 Methodology

Convolutional Neural Network based object detection architectures have evolved over time. Some of the earlier ones consist of R-CNN, which uses a selective search to detect regions of interest (ROI) and CNN to classify them. With time, improvements over R-CNN were made and models like Faster R-CNN, etc. were developed. Nowadays, YOLO is another popular family of image detection models, which we will use in our study as well. For now, we will use two models: Faster-RCNN and EfficientDet.

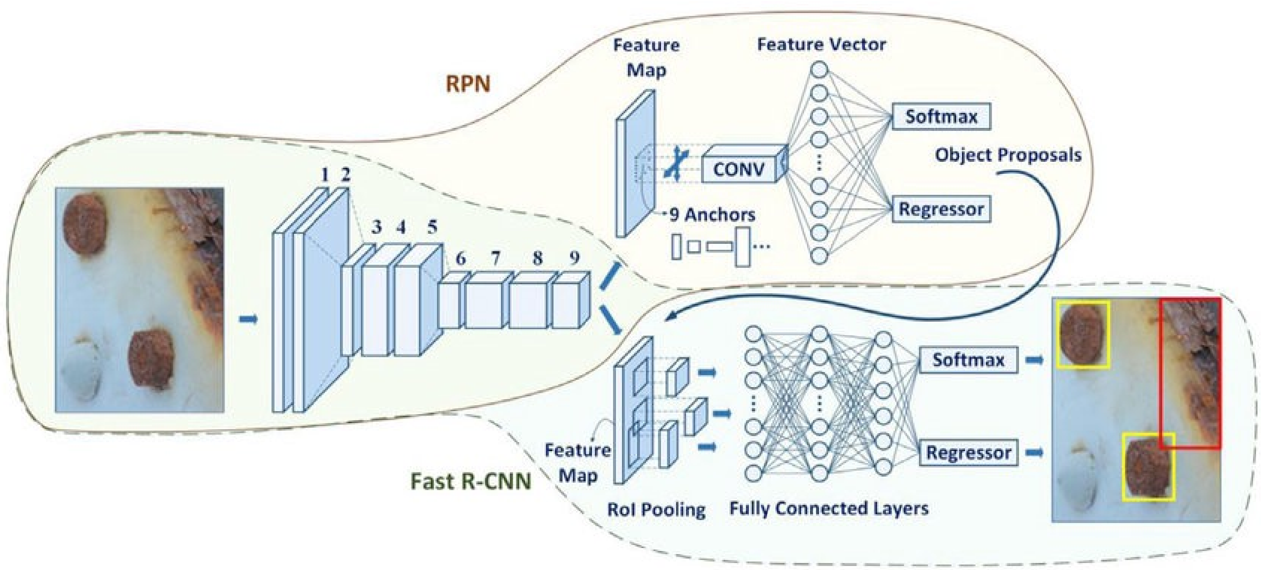


Fig 3.2.1 A brief look at the Faster R-CNN (Fast R-CNN + RPN) model (courtesy: [Google Images](#)).
The region proposal network takes a convolutional feature map that is generated by the backbone layer as input and outputs the anchors generated by sliding window convolution applied on the input feature map, these outputs are then sent for ROI pooling and then fed to a FCN for the final output.

The models were implemented using Python packages like Keras and PyTorch framework for training purposes. The experimental environment is provided by Kaggle, with the following system specifications: Intel Xeon (R) CPU, 2.54 GHz Processor, 12 GB system memory (RAM), NVIDIA GPU Tesla T4 (16GB VRAM).

For the Faster-RCNN model, the training is carried out with a batch size of 4 with the Stochastic Gradient Descent optimizer. The learning rate is initially set to 0.0001. The complete network is trained for 2 epochs (650*2 = 1300 iterations) and the weights were saved here. The training in total took about 35 minutes.

For the EfficientDet model, the training is carried out with a batch size of 2 with the Stochastic Gradient Descent optimizer. The learning rate is initially set to 0.0002, and then a plateau based scheduler is used to decrease the learning rate as the training progresses.

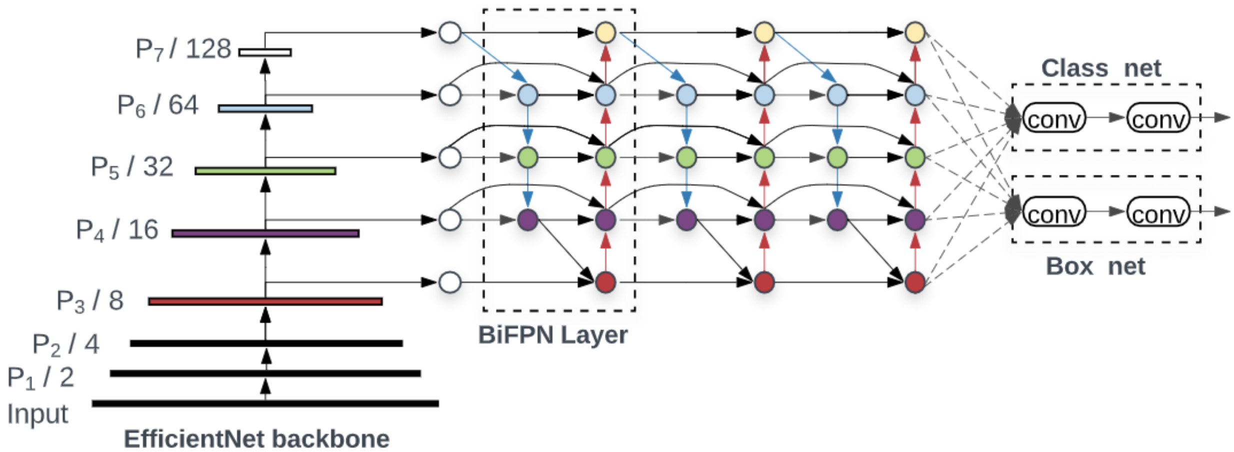
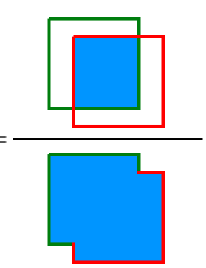


Fig 3.2.2 EfficientDet architecture visualised (courtesy: [Google Images]).
EfficientDet utilizes several optimization and backbone tweaks, such as the use of a BiFPN, and a compound scaling method that uniformly scales the resolution,depth and width for all backbones, feature networks and box/class prediction networks at the same time.

Data augmentation (steps like random cropping & flipping, changing brightness, contrast, hue values) is performed before feeding the images to the network in order to increase the robustness of the training. The complete network is trained for 4 epochs (1300*4 ~ 5200 iterations) and the weights were saved here. The training roughly took about 1 hour.

For the evaluation of both models, we use the Intersection over Union (IoU) metric as our primary detection evaluation metric. It is calculated as follows:

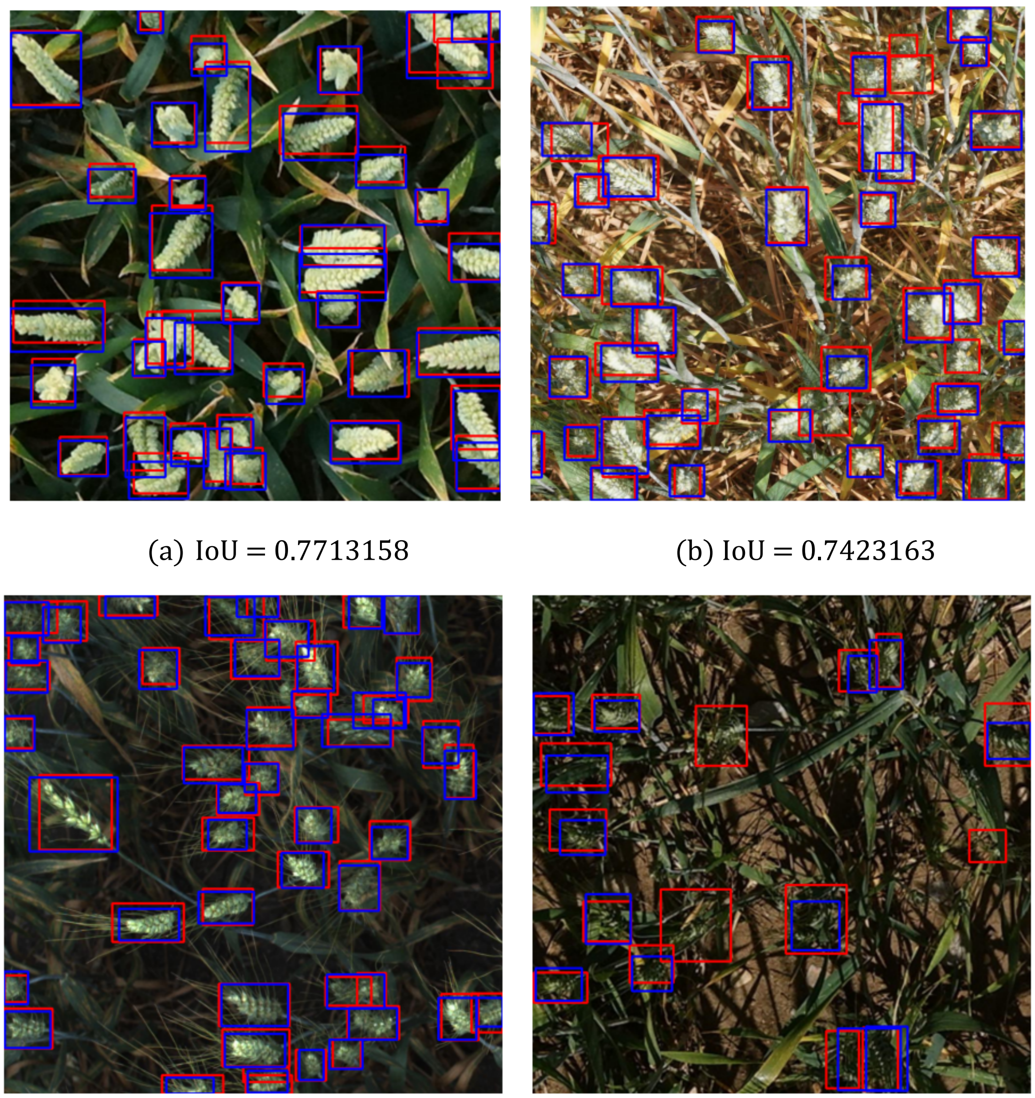
$$IOU = \frac{\text{area of overlap}}{\text{area of union}} =$$



Non-max suppression technique was followed, with an IOU threshold of 0.5, in order to remove the redundant predictions, which can affect our final evaluation metric.

4. Results and Discussion

For the Faster RCNN Model:



(a) IoU = 0.7713158 (b) IoU = 0.7423163

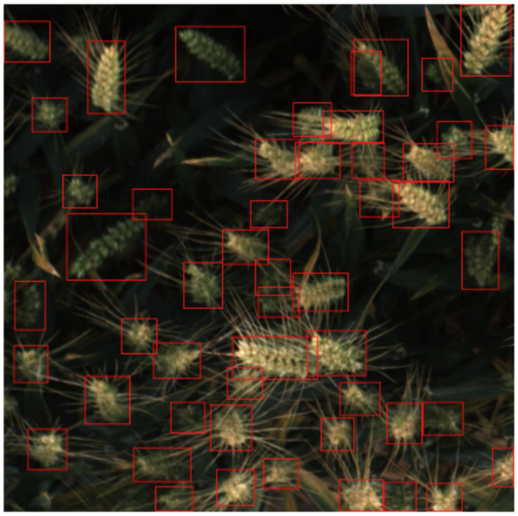
(c) IoU = 0.7676855 (d) IoU = 0.6195816

Fig 4.1 Bounding Boxes predicted by the Faster RCNN model on the test set images

Red boxes represent the ground truth, whereas the blue boxes represent the model predictions over test set images

We obtained an IOU of 0.6869 over the test set, and 0.7524 over the training set using the Faster-RCNN model.

For the EfficientDet model:



(a) Score = 0.784

(b) Score = 0.641

Fig 4.2 Bounding Boxes predicted by the EfficientDet model on the test set images

We obtained an IOU of 0.7066 over the test set, and 0.7464 over the training set using the EfficientDet model.

	Confidence Score	**Intersection over Union (IOU)**
Faster RCNN	0.9290	0.6869
EfficientDet	0.7622	0.7066

Table 4.1 Evaluation metrics of the models on the test dataset

5. Plan of Work

First Semester

5.1	Identified the important attributes of the Wheat Head Detection from the GWHD Dataset	Completed
5.2	Developed a Faster-RCNN and EfficientDet model for the detection of Wheat heads and yield estimation.	Completed
5.3	Compared the performance of developed models for yield estimation in terms of IOU, confidence score and computational time.	Completed

Second Semester

5.4	Development and evaluation of other detection models like YOLO, Mask-RCNN, etc.	In Progress
5.5	Implementation of other metric functions like Mean Average Precision (mAP) in order to increase the quality of benchmark.	Remaining
5.6	Training models from other wheat head data sources like ACID dataset, SPIKE dataset, etc.	Remaining

6. References

[1] Etienne David, Simon Madec, Pouria Sadeghi-Tehran, Helge Aasen, Bangyou Zheng, Shouyang Liu, Norbert Kirchgessner, Goro Ishikawa, Koichi Nagasawa, Minhajul A Badhon, et al. *Global wheat head detection (GWHD) dataset: a large and diverse dataset of high-resolution rgb-labelled images to develop and benchmark wheat head detection methods. Plant Phenomics, 2020.*

[2] Sandesh Bhagat, Manesh Kokare , Vineet Haswani, Praful Hambarde, Ravi Kamble, et al. *WheatNet-Lite: A Novel lightweight Network for Wheat Head Detection*

[3] Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun, et al. *Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks*

[4] Drew J. Lyon, Robert N. Klein, et al. *Estimating Winter Wheat Grain Yields.* University of Nebraska–Lincoln

[5] Mingxing Tan Ruoming Pang Quoc V. Le Google Research, Brain Team, *EfficientDet: Scalable and Efficient Object Detection*

[6] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi, *You Only Look Once: Unified, Real-Time Object Detection*

[7] Zhong-Qiu Zhao, Member, IEEE, Peng Zheng, Shou-tao Xu, and Xindong Wu; *Object Detection with Deep Learning: A Review*

[8] Non Maximum Suppression, https://paperswithcode.com/method/non-maximum-suppression