Wheat Head Detection using Image Analysis and Deep Learning

Ardhendu Singha

July 2020

1 Introduction

Wheat is one of the main crops grown around the world and is a part of our daily nutrition. During the Doubly Green Revolution[1] in 1950's, new wheat growing techniques ensured food security in the emerging countries; saving thousands of people from starvation. However last few decades have seen stagnation in the rate of increase in wheat yields. Even though phenotype research is giving us numerous data for genetic gain, lack of new techniques which can yield high amount of growth is still posing a threat. For farmers, assessing the production of wheat heads, making vital decisions about crop yields and detecting diseases in early stages in of paramount importance. Thus it is becoming a challenge to develop highly accurate and robust models to extract information form wheat images.

In recent years we have seen monumental advancements in the fields of Machine Learning. Increments in GPU performances and emergence of large data sets has enabled us to create and validate robust Deep Learning models; which are essential for wheat head detection. Several methods have been proposed for wheat head quantification from RGB high-resolution images. In [2] the authors demonstrated the potential to detect wheat heads with a Faster-RCNN[3] object detection network. But these detection methods are trained on individual limited data sets, which makes it difficult to extrapolate information from new examples. Another thing to worry about is that, the images of a wheat field is by nature messy and have numerous variability condition which can make it difficult to get a clear photo. Again we also have to consider different weather conditions, winds, maturity of the crop (mature golden ones can mix with the color of the ground and younger green heads camouflage themselves with the stem) and availability of infrastructures to fulfill all the requirements needed to make a quality data sets. Further, because the labelling process is burdensome and tedious, only a small fraction of the acquired images are processed. Finally, labelling protocols may be different between institutions, which will limit model performance when trained over shared labelled data sets. To fill the need for a large and diverse wheat head data set with consistent labelling, Global Wheat Head Detection(GWHD) data set ahs been developed, that can

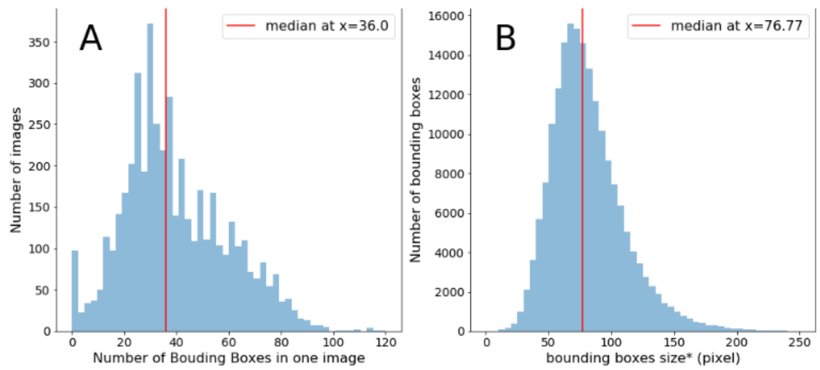

Figure 1: Distribution of the number of bounding boxes per image (a) and bounding boxes size* (b) in the GWHD Dataset.

be used to benchmark methods proposed in the computer vision community. The GWHD dataset results from the harmonization of several datasets coming from nine different institutions across seven countries and three continents. This makes it a very unique and proper dataset to validate models and judge their robustness.

2 Dataset Description

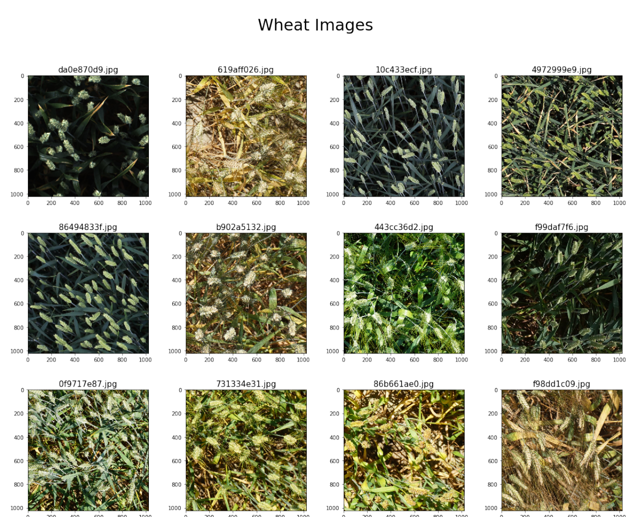

Figure 2: Sample Images

The GWHD dataset represents 4,698 squared patches extracted from the

2219 original high-resolution RGB images acquired across the 11 sub-dataset. It represents 188,445 labelled heads which average 40 heads per image in good agreement with the 20 to 60 targeted heads per image [Figure 1(a)]. The size of the bounding boxes around the heads shows a slightly skewed Gaussian distribution with a median typical dimension of 77 pixels [Figure 1(b)].

The diversity of acquisition conditions sampled by the GWHD dataset is well illustrated in Figure 3: illumination conditions are variable, with a wide range of headsand background appearance. Further, we observe variability in head orientation and view directions, from an almost nadir direction up to a mostly oblique direction as in the case of ETHZ-1 (Figure 3). But GWHD

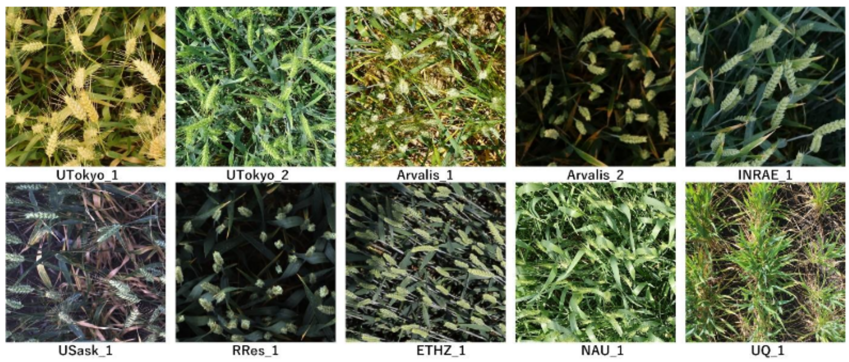

Figure 3: Example of images from different acquisition sites after cropping and rescaling.

Figure 4: Example of images with 10 boxes or less.

poses some messy images. It includs about 100 images that contain no heads to represent in-field capturing conditions and add difficulty for benchmarking. Few images contain more than 100 heads with a maximum of 120 heads. But this makes developing a generic model all the more difficult. Let's look at some of the examples.

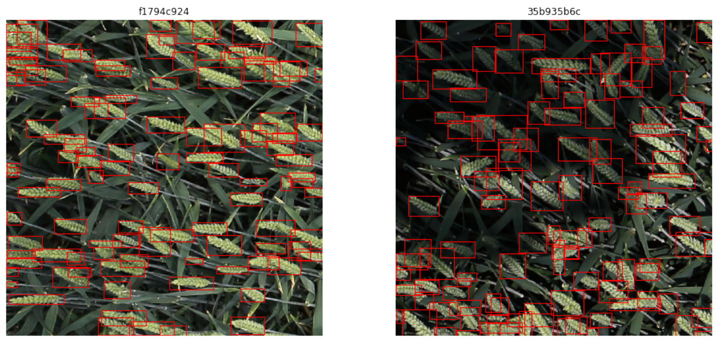

Figure 5: Example of images with 100 boxes or more.

3 Model Building

3.1 Data Augmentation

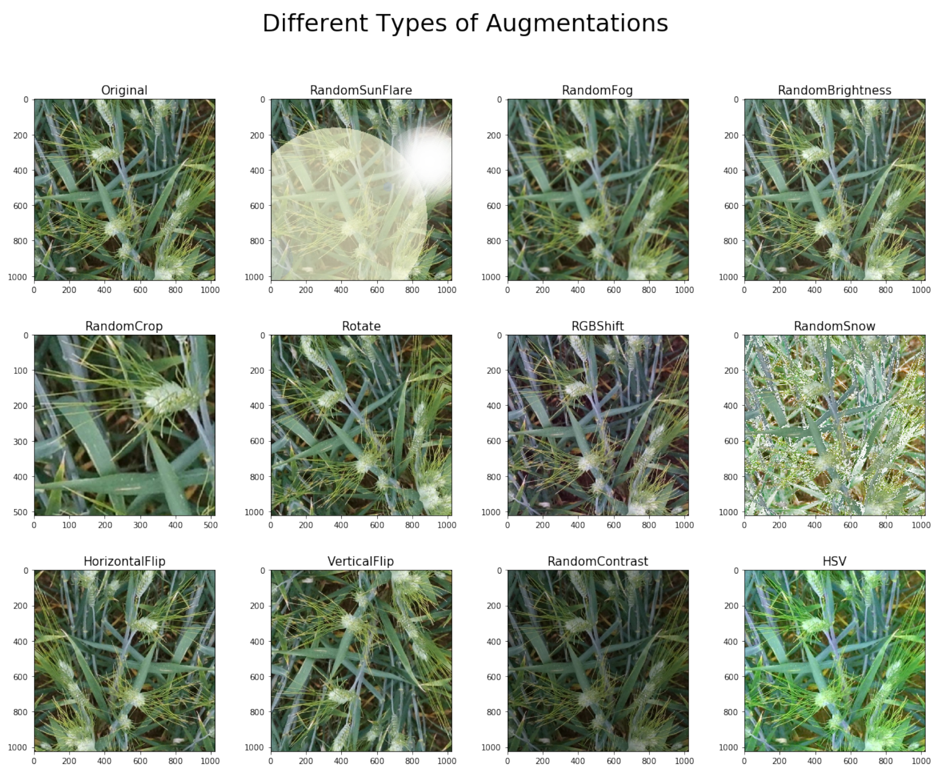

Figure 6: Types of augmentation used.

There are 3422 training images and only 10 are available for testing. For making the model work we need augment the images. Albumation provides us with simple yet powerful tools to transform the images during training and testing with integrated support for pytorch and tensorflow. We will make use of that here.

3.2 Evaluation Metrics

GWHD dataset is labelled by bounding boxes indicating a wheat head in the foreground which allows us to apply object detection methods. We are going to use the standard evaluation metrics such as mean average precision computed from true and false positives to judge performance of the model. A true positive corresponds to a predicted bounding box with an intersection over union (IoU) greater than and equal to 0.5 with the closest labelled bounding box(Ground Truth box). A false positive corresponds to a predicted bounding box with an IoU strictly lower than 0.5 with the closest labelled bounding box.In the case of two predicted boxes with an IoU greater than or equal to 0.5 on the same bounding box, the most confident one is considered as a true positive and the other as a false positive. And the final mean Average Precision noted as mAP@0.5 is the considered metric for evaluating the localization performance of the model. Given predicted box: A and a ground truth box: B with threshold: t following is our metrics,

$$IoU = \frac{A \cap B}{A \cup B} \tag{1}$$

$$Prescision(t) = \frac{TP(t)}{TP(t) + Fp(t) + FN(t)} \tag{2}$$

$$mAP = \sum_{t} Precision(t) \tag{3}$$

3.3 Training Model

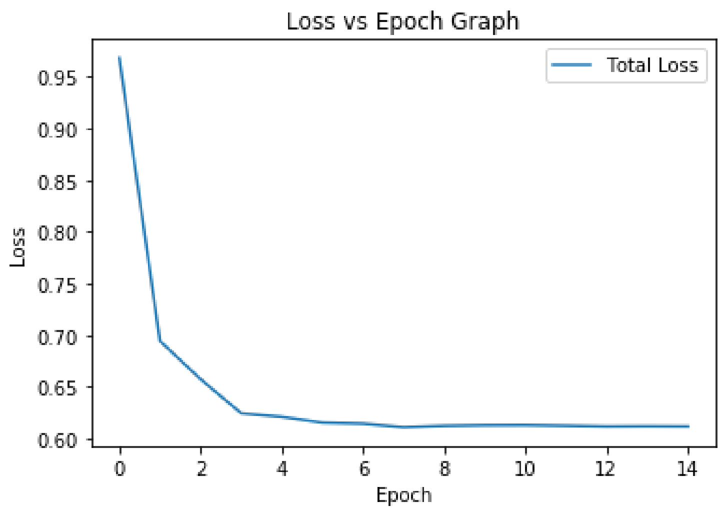

Figure 7: Loss vs Epoch graph

For training the GWHD dataset the most popular choice would be Faster-RCNN with a ResNet50 as the backbone. PyTorch provides us with dataloaders which acts as a pipeline and automatically loads the the data into the model. The model is trained for 15 epoch with batch sizes of 8. The data is augmented to avoid overfitting and original size of the images are kept. As for the optimizer, I have used Stochastic Gradient Descent with learning rate: 0.005, momentum: 0.9 and weight decay: 0.0005. COCO evaluate has been used to evaluate the model with a validation dataset. After Epoch15 we have mAP@0.5 of 0.880 and loss of 0.582.

3.4 Results

We have 10 test images available. In Figure 8 we can see that majority of the wheat heads are detected accurately but some of the bounding boxes are overlapping which indicates that that their is bias and the model is confusing between choosing the right sizes of the prediction boxes.

Figure 9 shows us the comparison between predicted boxes and ground truth boxes. Prediction boxes are really close to the ground truth boxes. The model is performing better here; that indicates that the model is not ready for a generic test yet. We have to find better ways to fix this.

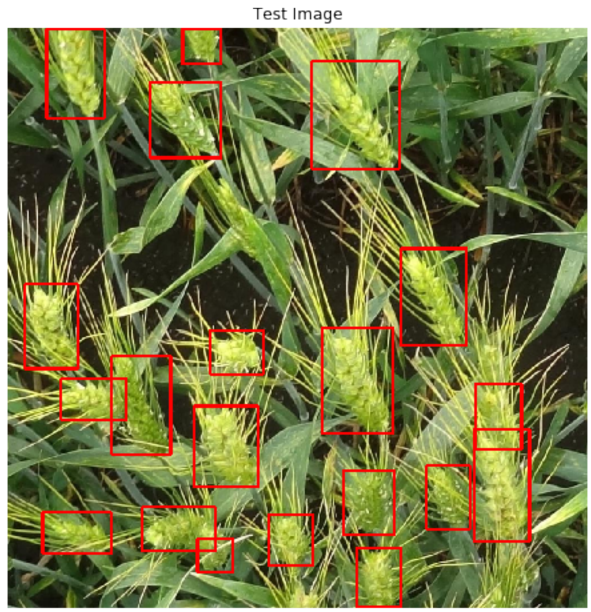

Figure 8: Testing example

Figure 9: Predicted VS Ground Truth

4 Conclusion

The results suggest that the performance of the model is still poor, though it had a mAP@=0.5 of 0.880 the mAP@0.5:0.75 is at 0.627; which indicates that the predicted bounding boxes have much less probability of detecting a wheat head. To overcome these difficulties we need more data; The backbone used here, ResNet50 is a deep network which will benefit from large dataset. We can use pseudo labelling techniques with soft Non-Maximum Suppression to gather more augmented data. The relatively poor performance of this standard object detection network on the GWHD dataset provides an opportunity for substantial future improvement.

References

[1] P. Sadeghi-Tehran H. Aasen B. Zheng S. Liu N. Kirchgessner G. Ishikawa K. Nagasawa M.A.Badhon C. Pozniak B. de Solan A. Hund4 S.C. Chapman F. Baret I. Stavness W. Guo E. David, S. Madec. 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. 2020.

[2] H. Lu-B. De Solan S. Liu F. Duyme E. Heritier S. Madec, X. Jin and F. Baret. "ear density estimation from high resolution rgb imagery using deep learning technique". *Agric. For. Meteorol.*, 264:225–234, Jan. 2019.

[3] Ross Girshick-Jian Sun Shaoqing Ren, Kaiming He. Faster r-cnn: Towards real-time object detection with region proposal networks.