HEALTHY VS UNHEALTHY MAIZE CORN OBJECT DETECTION WITH DEEP LEARNING-BASED CLASSIFICATION ON MULTI-SPECTRAL IMAGES

Apolline Blachet[1] (s222903) and Jónas Már Kristjánsson[1] (s223596)

1 DTU Compute, Technical University of Denmark

Introduction

This project addresses the challenge of detecting and classifying maize in multi-spectral images, responding to the quality control requirements of the food industry for precise identification of maize seeds. While traditional approaches rely on image processing and machine learning, this project delves into the potential of deep learning for seed detection and classification. The aim is to locate and classify each maize corn to be either healthy or unhealthy maize.

Our solution is a object detection pipeline consisting of selective search for object localization, VGG-16 based CNN for classification, and non-maximum suppression for post-processing.

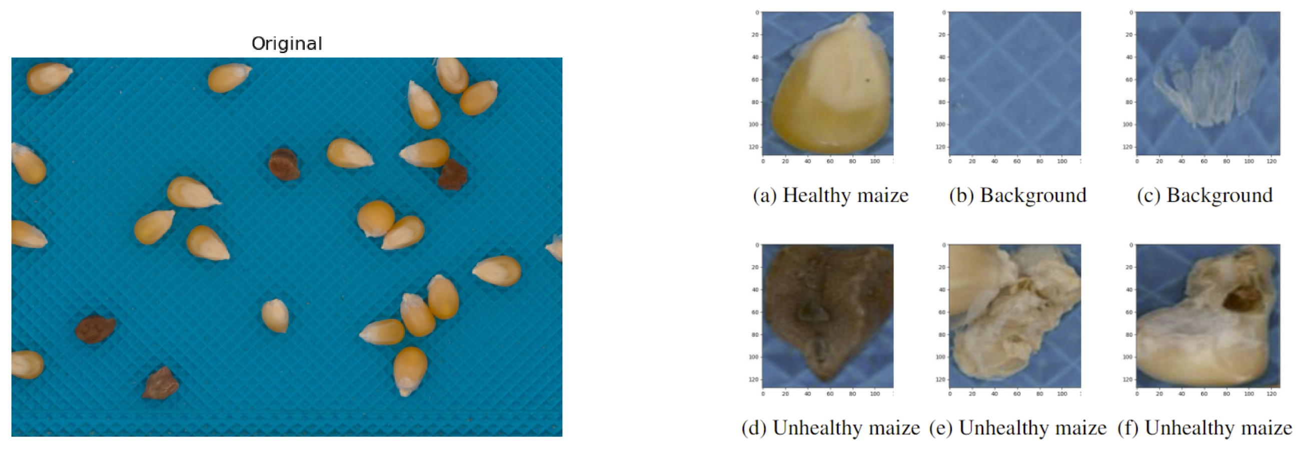
Data set

The data set was provided by Videometer A/S [1]. Consist of

- 10-band Multi-Spectral Images
  - 1200 x 1758 pixels
  - Covering a spectrum from Ultra-violet (365nm) to Near-Infrared (970nm)
- 129 images for training and 45 images for testing
- With 3 labelled ground truth bounding boxes, Background - Maize Healthy - Maize Unhealthy.

For the classifier image patches for each label was located in the training images.

- 688 image patches for each label, totalling 2064 image patches
- Divided in training 75%, validation 15% and testing 10%
- Variable sizes from 40-400 pixels each side. Resized to 128 x 128
- Randomly rotated between -179 and 179 degrees



Example of sRGB transformed image

Example of labelled image patches

Figure 1: Examples from the data set

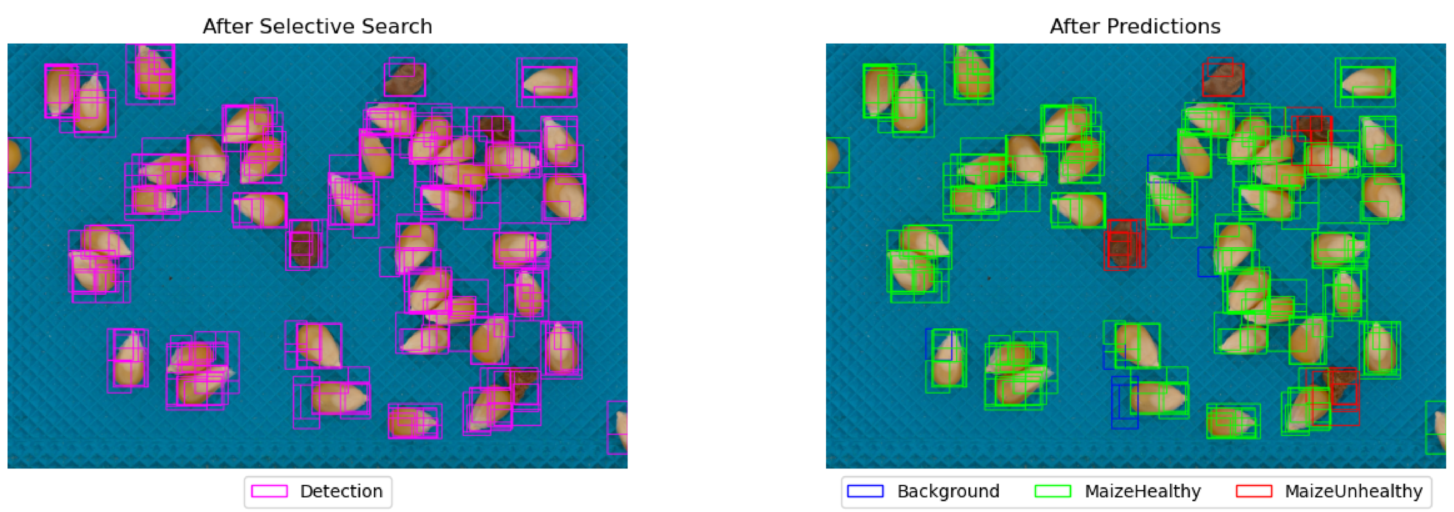
Selective Search

Selective search is a object of interest location suggestion algorithm introduced in [3]. It utilizes the proposed hierarchical group algorithm to combine object location suggestions of several variations. Using a single level of color (HSV), fill, texture and size strategy described in [3] is applied to the sRGB transformed multi spectral images. The strategy was initalized with k=200 which sets a scale of observations and Gaussian derivation with σ=0.8 gave the most promising results, then filtered out all bounding boxes with either width or height outside of 3 to 10**%** of the image width. Example of results can be seen in fig 2a.

Classifier

The VGG-16 model was used as a base and modified so the input layer can take in 10 channel images and output of size 3 with a softmax activation. Initializing the weights with the pre-trained weights from the ImageNet competition [2] and training the entire model on the image patches data set with the following

- Cross-Entropy Loss
- Adam Optimizer with a learning rate of 10^{-6}
- Batch size of 32
- Number of epochs as 14



(a) Detected objects marked in pink

(b) Detected objects with predicted labels

Figure 2: Object localization and classification results

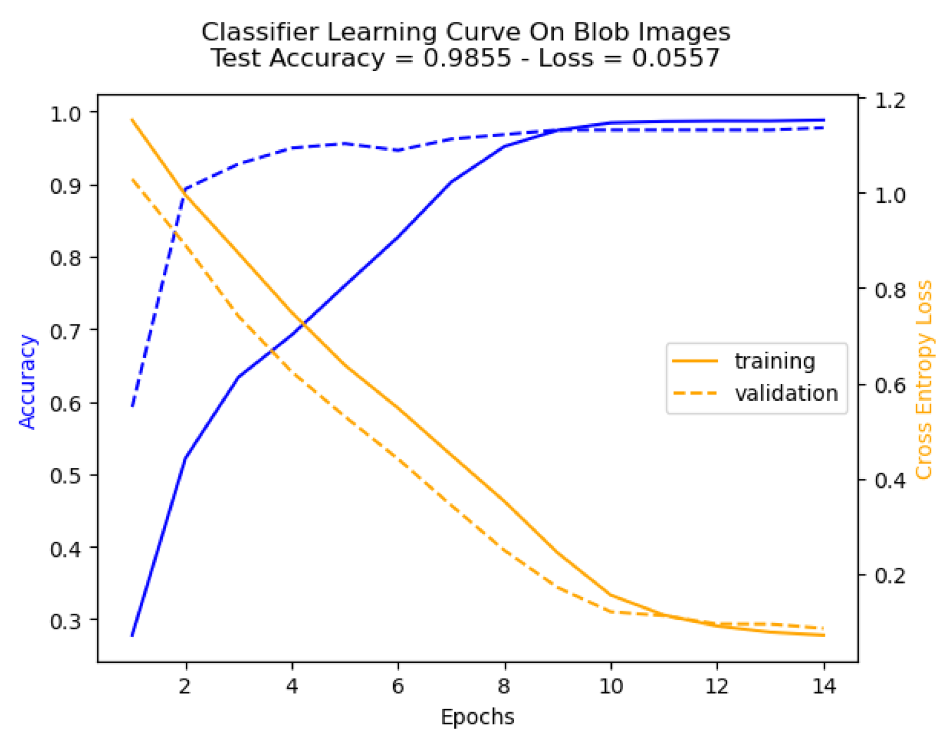


Figure 3: Learning curve of the classifier

Non-Maximum Suppression

To have fewer predictions of the same object we merge bounding boxes that are assumed to belong to the same object utilizing the GreedyNMS algorithm. Which greedily selects higher confidence scoring detections over close-by less confident neighbours. The distance used to compare detections distances is the Intersect-Over-Union (IoU). For simplification the confidence score threshold was set to 0.8. As can be seen in 4 the optimal value of IoU threshold is 0.1 resulting in $\text{mAP}^{IoU=0.5} = 0.8997$ and best precision-recall curve on the training data. Using the same hyper-parameters on the testing data the $\text{AP}^{IoU=0.5}$ for the healthy maize label was 0.8985 and for the unhealthy maize label was 0.9491 resulting in a $\text{mAP}^{IoU=0.5} = 0.9238$ as can be seen in 5.

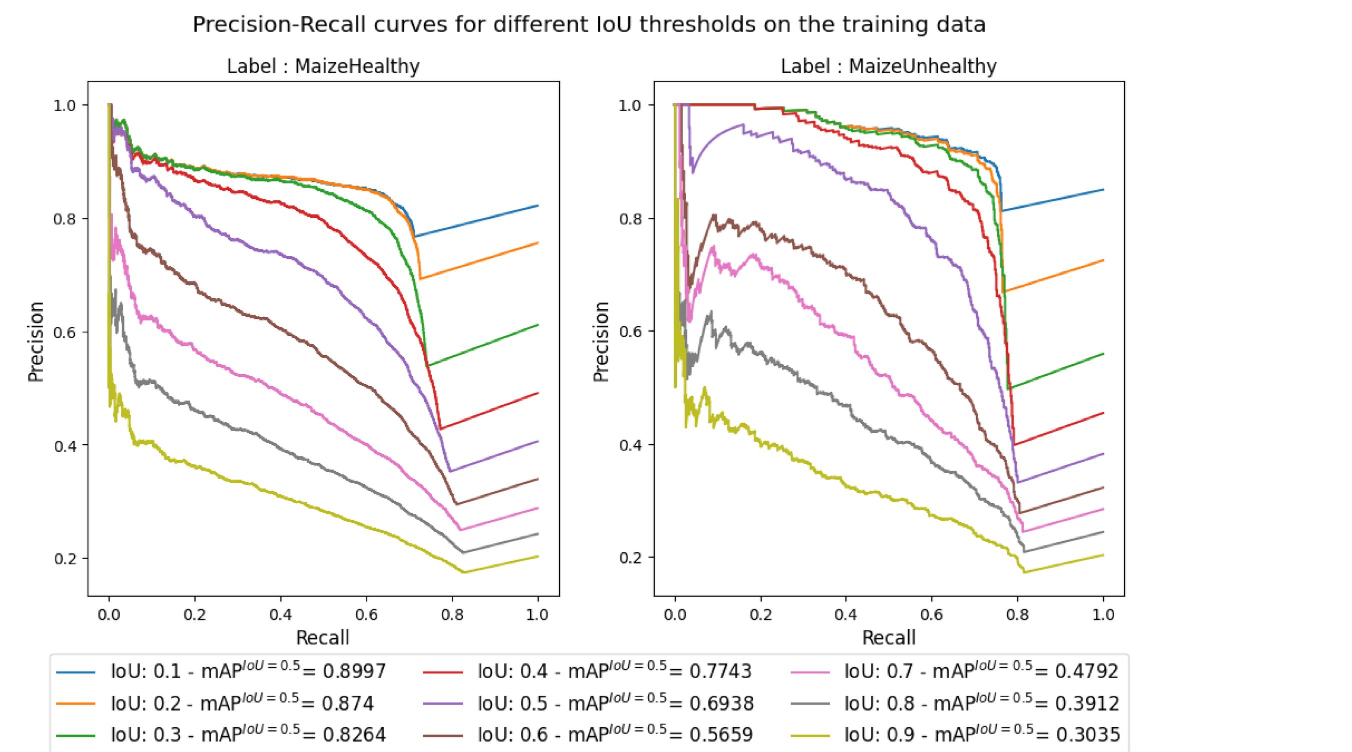


Figure 4: Precision-Recall curve along $\text{mAP}^{IoU=0.5}$ value for each IoU threshold on the training data

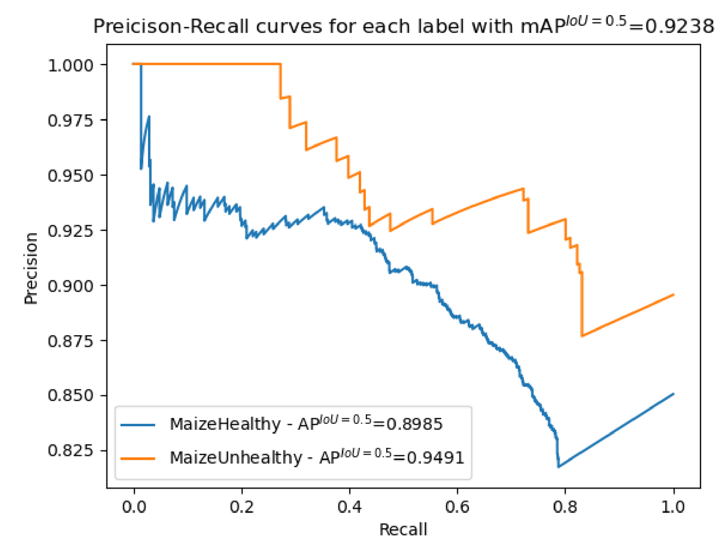
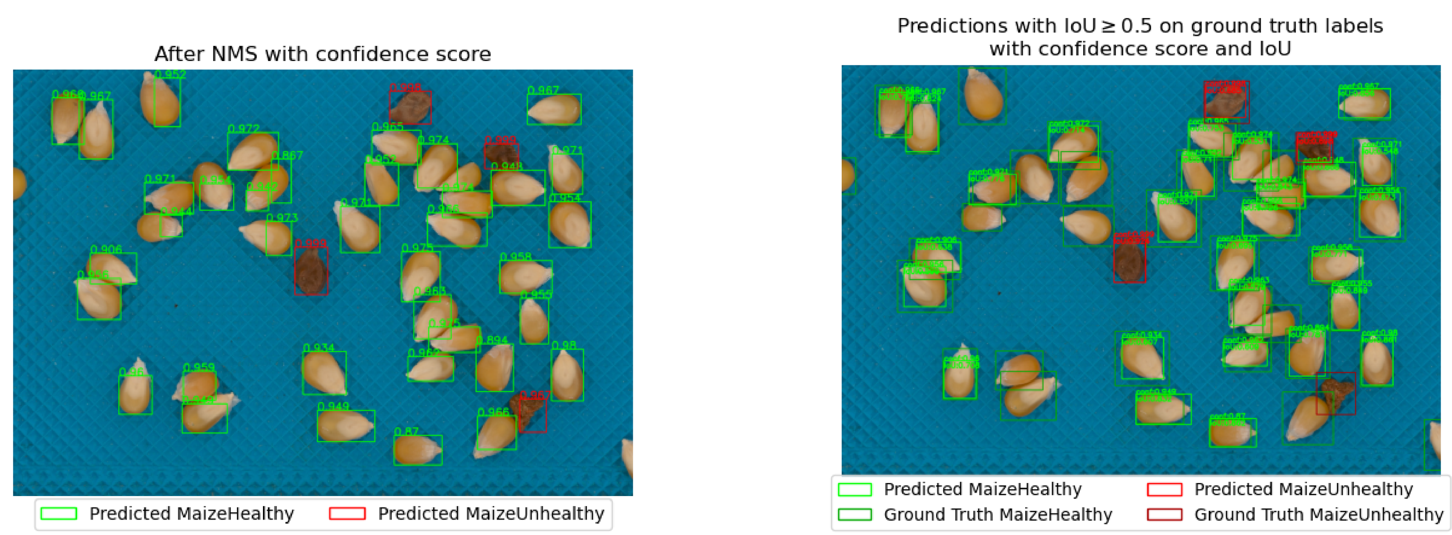


Figure 5: Precision-Recall curve along with $\text{mAP}^{IoU=0.5}$ value for the testing data



(a) Results after NMS with confidence score

(b) Final comparison with confidence score and IoU

Figure 6: Example of detections after NMS and then final comparison

Discussion

The classifier performed well but with only 14 epochs to hit the training ceiling indicates overfitting but that doesn't seem to be the case with a test accuracy of 0.9855.

The Selective Search separated most adjacent objects with the same label, as can be seen on the left side in fig 2a. But since the optimal IoU is low then it encounters difficulty when many objects are clustered as can be seen in in the top right of 6a. This could be improved by increasing the IoU threshold, but then the algorithm would generate significantly increased number of false positive predictions as can be seen in fig 4. This is mostly due to multiple same label predictions of the same object or one predictions for multiple ground truths, thus the goal for the improvement would be to increase the quality of the object location detection. This could be done on a few different ways. The first one is to improve the tuning of the hyper-parameters f.x. lowering the k threshold which would favor larger objects. This would introduce other problems like smaller ones not being detect or a group of corn being detected as one. The second suggested improvement would be to add more layers to the selective search. This would definitely help and improve the suggestions but are more computationally expensive which would effect the real-time performance. The last suggested improvement is to use another object localization algorithm which would be more computationally efficient or more specific to this application. This is thought to yield the optimal solution as the background is standardized both in texture and color, as well as the camera.

This feasibility test was successful since this method showed real potential of providing robust and fast results. For future work it would be interesting to try this pipeline on broader data set of maize or other kernels or seeds to see assess the replicability of the results.

Acknowledgements

The authors wish to thank Jens Michael Carstensen and Videometer A/S for the data set.

References

[1] V. A/S. Company documentation. URL `https://www.videometer.com`.

[2] Pytorch. Vgg documentation. URL `https://pytorch.org/vision/main/models/vgg.html`.

[3] 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*, 104(2):154–171, 2013. URL `https://ivi.fnwi.uva.nl/isis/publications/2013/UijlingsIJCV2013`.