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

Apolline Blachet (s222903), Jónas Már Kristjánsson (s223596)

Technical University of Denmark

ABSTRACT

The report presents a deep learning-based approach for maize detection and classification in multi-spectral images. The methodology involves selective search for object localization, VGG-16 based Convolutional-Neural-Network for classification, and non-maximum suppression for post-processing. The dataset, provided by Videometer, includes healthy and unhealthy maize seeds captured in multi-spectral images and ground truth labels.

The selective search algorithm successfully identifies potential objects, but challenges arise with clustered objects. Fine-tuning hyperparameters or exploring alternative algorithms is suggested for improvement. The VGG-16 based classifier achieves a test accuracy of 98.55% in training, showcasing its effectiveness in classifying healthy and unhealthy maize seeds.

Non-maximum suppression (NMS) is employed to refine bounding box predictions, resulting in a full model mean average precision (mAP) of 92.38% on the testing data. Precision-recall curves demonstrate the trade-off between precision and recall for different IoU thresholds.

The results indicate the feasibility of the proposed approach for maize detection and classification in multi-spectral images. Suggestions for further improvement include optimizing selective search parameters and exploring alternative object localization algorithms. Overall, the system shows promise for real-world applications in agriculture and food industry quality control. The full code can be found in `https://github.com/jonasmark97/DeepLearning_Final_Project`.

***Index Terms*—** Multi-Spectral Images, Maize, Classification, Object detection, VGG-16, Transfer Learning, Selective Search, Non-maximum Suppression.

1. 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 dataset, provided by Videometer, a company specializing in multi-spectral imaging, comprised of 10-band multi-spectral images capturing both healthy and unhealthy maize seeds against a blue background. The complexity arises from the diverse forms of unhealthy maize and potential confounding factors in the background, such as small particles and maize skin. The report details the methodology involving selective search for object localization, VGG-16 based CNN for classification, and non-maximum suppression for post-processing. The full code can be found in `https://github.com/jonasmark97/DeepLearning_Final_Project`.

2. DATA

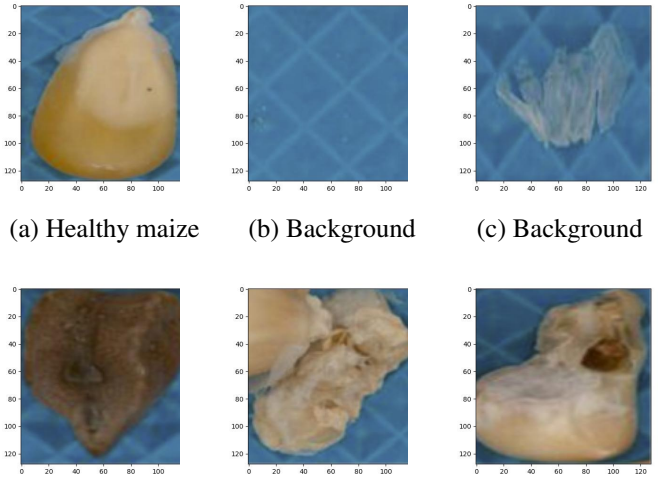
Our dataset, provided by Videometer A/S [1], consists of multiple 10-band multi-spectral images. This dataset is divided into two subsets: 129 images for training and 45 images for testing. The images showcase various objects set against a blue background, specifically healthy maize, which is viable for crop yield, and unhealthy maize, exhibiting diverse manifestations like dark and desiccated kernels, split kernels, or colorless kernels. Unlabeled areas within the images represent the background. Approximately 10 percent of all seeds in the dataset are unhealthy maize. Each image is sized 1200x1758 pixels, covering a spectrum from Ultra-violet (365nm) to Near-Infrared (970nm). Refer to Figure 1 for an RGB visualization of a typical image.

The objective is to conduct object detection and classification on these images. For each image, the ground truth includes bounding boxes delineating the approximate position of each seed and associated labels indicating whether it is healthy or unhealthy maize. Figure 10 provides a visual example of this ground truth.

Additionally, we have a separate dataset used for training the classification model. This dataset comprises 2064 images of varying sizes from 40-400 pixels on each side , depicting either a single healthy or unhealthy maize seed or background. With 688 images for each category, the seeds in these images are aligned with their length directions. Examples of these images are displayed in Figure 2. It's noteworthy that

Fig. 1: RGB visualization of a typical image from the dataset

unhealthy maize exhibits various appearances, and the background may include maize skin and residuals, which should not be identified as objects within the scope of our project. This supplementary dataset is essential for training the classification model, ensuring a balanced representation of background, healthy, and unhealthy maize images.



(a) Healthy maize (b) Background (c) Background

(d) Unhealthy maize (e) Unhealthy maize (f) Unhealthy maize

Fig. 2: Visualization of typical non-rotated images used for training the classification model

3. METHODOLOGY

3.1. Selective search

Selective search is a object of interest location suggestion algorithm introduced in [2]. It utilizes their proposed hierarchical group algorithm to combine object location suggestions of several variations. The implementation used in this report is from openCV [3], where their Single Strategy is used which is the simplest version of Selective Search and thus the fastest. Using only the single level of color (HSV), fill, texture and size strategy described in [2] is applied to the sRGB transformed multi spectral images implemented in Videometer python toolbox [4].

After the selective search the suggested bounding boxes are filtered by height and width. The sizes of objects of interest are roughly 5 to 8% of the width and fairly equal on each side, but we added a 2% safety margin. Then removed all bounding boxes with either width or height outside of 3 to 10% of the image width.

3.2. Classification Model

The objective is to construct a model that takes the extracted patches from the multi-spectral images and classifies into one of three categories: healthy maize, unhealthy maize or background.

The chosen architecture for this task is the VGG-16 model, a deep convolutional neural network renowned for its simplicity and effectiveness in image classification tasks. Figure 3 provides a schematic view of its architecture. Notably, VGG-16 was originally trained on the ImageNet dataset, which comprises 14 million RGB images categorized into 1000 classes, such as ”balloon” or ”car.” [5]

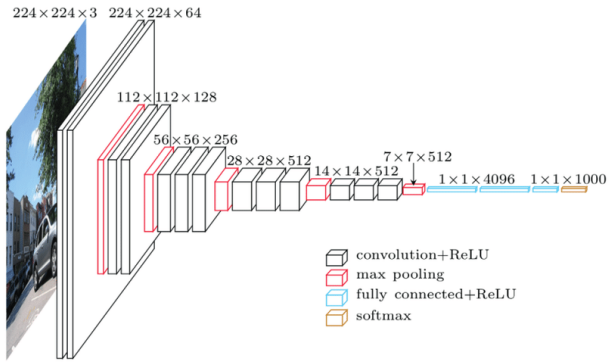


Fig. 3: VGG-16 architecture (Ref. [6])

For our project, modifications were made to the input layer, enabling the model to process 10-band multi-spectral images instead of RGB images. Additionally, adjustments were implemented in the output layer to classify inputs into the specified three categories, deviating from the original 1000 classes. Along with an softmax output layers activation.

To address our specific classification problem, the model was trained using the 2064 multi-spectral images, examples of which are shown in Figure 2. The dataset was divided into subsets for training (75%), validation (15%), and testing (10%). Each image resized to 128 by 128. Given the potential

rotations between -179 to 179 degrees of seeds in the global images, each of the 2064 images underwent random rotations as a data augmentation technique. Subsequently, the entire model was retrained, initializing the weights with those of the original VGG-16 model pre-trained on the ImageNet dataset.

3.3. Merging bounding boxes

To have fewer predictions of the same object we merge bounding boxes that are assumed to belong to the same object. Often refered to as Non-maximum suppression (NMS) and will be here after. Utilizing the implementation from openCV [3], GreedyNMS , who greedily selects higher confidence scoring detections over close-by less confident neighbours since they are likely to cover the same object. To compare if two detections are close enough to each other the metric Intersect-Over-Union (IoU) or Jaccard distance is used, it describes how much the two detections are overlapping and is formulated the following :

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

Where A and B are individual bounding boxes.

3.4. Evaluation metrics

To evaluate detection the metrics precision and recall will be used. They are defined as follows

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives + False Positives}}$$
$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives + False Negatives}}$$

For each label the Average Precision (AP) will be calculated and a detection will only be true positive when it is over 0.5 IoU with the ground truth with a same label, calling it $\text{AP}^{IoU=0.5}$. Furthermore, taking the mean over all the labels is called mean average precision or $\text{mAP}^{IoU=0.5}$.

4. RESULTS

4.1. Selective Search

The threshold k=200 which sets a scale of observations and Gaussian derivation with σ=0.8 gave the most promising results, as can be seen in fig 4.

4.2. Classification Model

The classifier's learning curve is shown in Figure 5. These results are obtained using Cross-entropy loss, the Adam optimizer with a learning rate of 10^{-6}, a batch size of 32, and 14 epochs. Notably, a test accuracy of 98.55% is achieved, particularly noteworthy given the relatively small number of epochs.

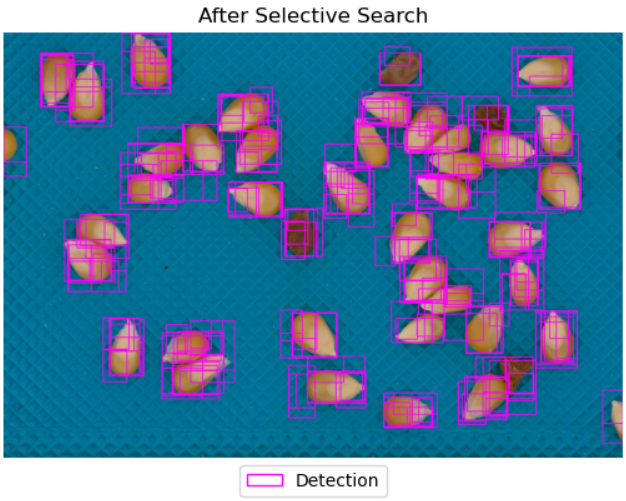


Fig. 4: Results after Selective Search, detections marked in pink

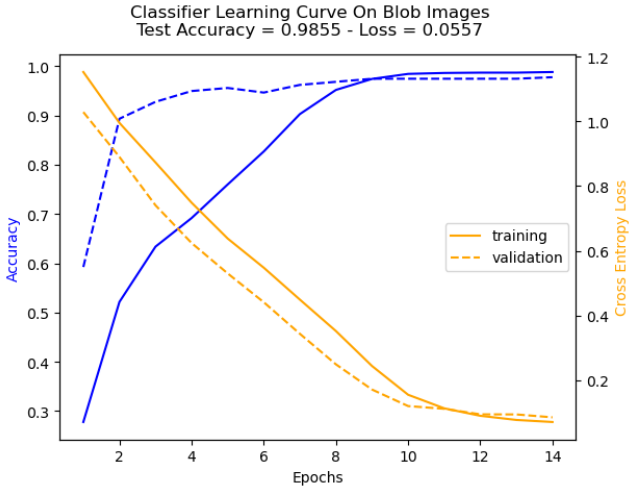


Fig. 5: Learning curve of the classifier

Once the model trained, it is used on the patches extracted from multi-spectral images using selective search (see Figure 4). The outcome for the respective image is illustrated in Figure 6, where each bounding box is assigned a predicted class.

4.3. Non-maximum Suppression

For simplification the confidence score threshold was set to 0.8. The optimal IoU threshold was found by running all the bounding box suggestion from the training data through the NMS algorithm with a range of IoU threshold from 0.1 to 0.9 with a step of 0.1. Then measuring the precision-recall curve when comparing it to the ground truth for each non-background label along with $\text{mAP}^{IoU=0.5}$ which can be seen

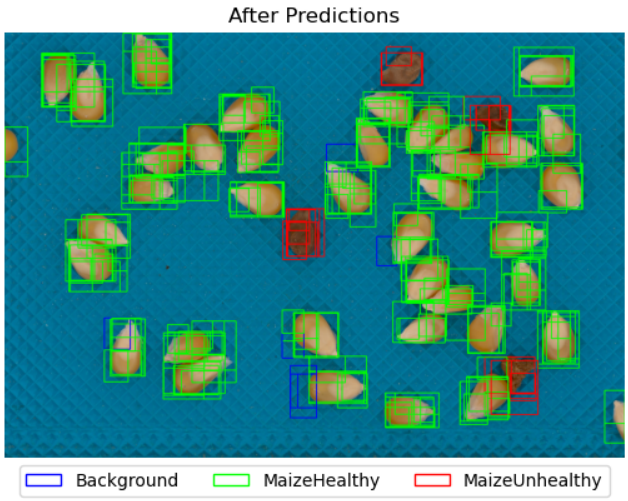


Fig. 6: Results after predictions. Healthy maize is marked in green, unhealthy maize in red, and background in blue.

in fig 7. As can be seen the optimal value of IoU threshold is 0.1 resulting in mAP$^{IoU=0.5}$ = 0.8997 and best precision-recall curve.

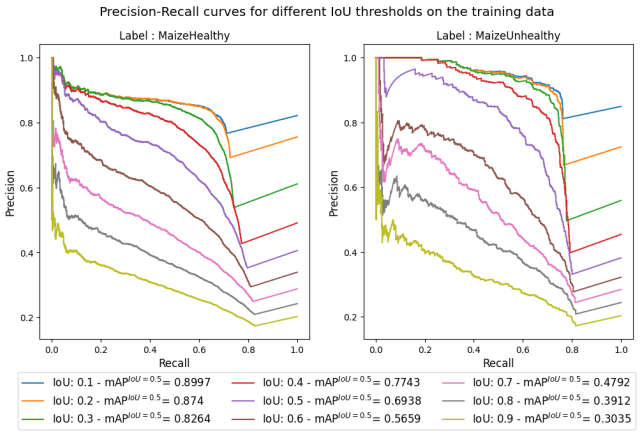


Fig. 7: Precision-Recall curve along mAP$^{IoU=0.5}$ value for each IoU threshold on the training data

Using a confidence score threshold of 0.9 and IoU threshold of 0.1 on the NMS the following precision-recall curve was calculated for the testing data in fig 8. The AP$^{IoU=0.5}$ for the healthy maize label was 0.8985 and for the unhealthy maize label was 0.9491 resulting in a mAP$^{IoU=0.5}$ = 0.9238 on the testing data.

5. DISCUSSION

The classifier performed well but with only 14 epochs to hit the training ceiling indicates overfitting but that doesn't seem

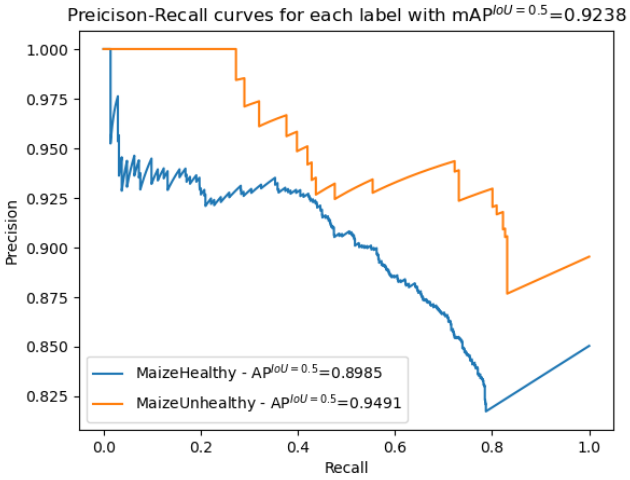


Fig. 8: Precision-Recall curve along with mAP$^{IoU=0.5}$ value for the testing data

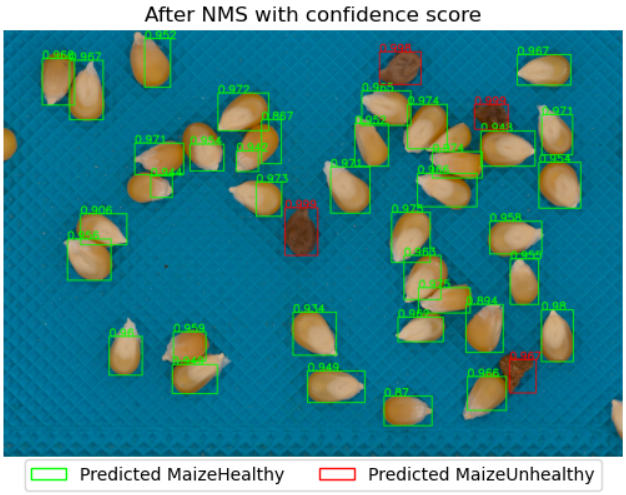


Fig. 9: Results after NMS with confidence score

to be the case with a test accuracy of 0.9855. The Selective Search did perform well and separated most adjacent objects with the same label, as can be seen on the left side in fig 4. 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 fig 9.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 7. 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, which all focus on improving the object suggestions. The first one is to improve the

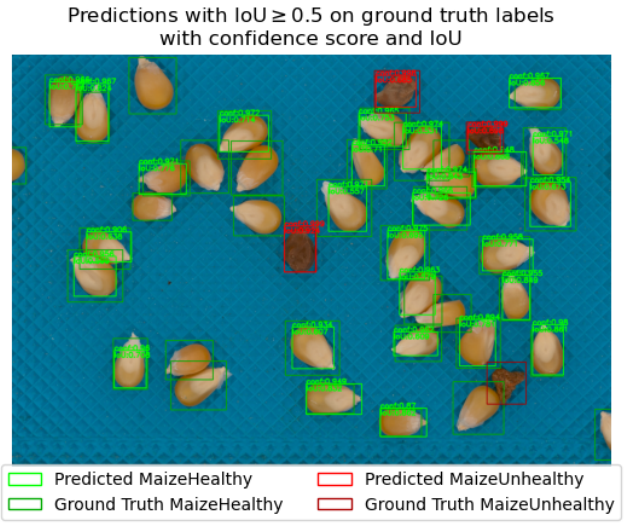


Fig. 10: Final comparison of detections and ground truth with confidence score and IoU

tuning of the hyper-parameters f.x. lowering the k threshold which would favor larger objects. This solution is not robust since this would be very specific for these sizes of corns i.e. if smaller ones would come through then it might not detect them or a group of corn being detected as one, but would certainly work on this data set. The second suggested improvement would be another strategy f.x. ”fast” or ”quality” in OpenCV implementation i.e. add more layers to the algorithm. 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 dataset of maize or other kernels or seeds to see assess the replicability of the results.

6. CONCLUSION

In this report the feasibility of deep learning-based approach for maize detection and classification in multi-spectral images was tested. Creating a classical object detection pipeline, comprised of three parts, object localization, classification and suppression. The object localization was performed using selective search and an example can be seen in 4. The classification was performed with a modified VGG-16 model initialized with weights from the ImageNet competition. It performed exceptionally well as can be seen in fig 5 with testing accuracy of 0.9855 and Cross-Entropy loss of 0.0557. The suppression was done with GreedyNMS and from the training data the optimal IoU threshold was found to be0.1 with a fixed confidence score threshold of 0.8, resulting in a $\text{mAP}^{IoU=0.5}$=0.8997, as can be seen in fig 7. Running the whole pipeline on the test dataset resulted in a $\text{mAP}^{IoU=0.5}$=0.9238, as can be seen in fig 8. Overall the selective search did separate adjacent objects and finding individual kernels but with the low IoU and GreedyNMS many of the true predictions were dropped. The feasibility test is concluded to be a success providing robust and fast results.

References

[1] V. A/S, “Company documentation.” [Online]. Available: https://www.videometer.com

[2] 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, no. 2, pp. 154–171, 2013. [Online]. Available: https://ivi.fnwi.uva.nl/isis/publications/2013/UijlingsIJCV2013

[3] OpenCV, “Selective search implementation from opencv.” [Online]. Available: https://github.com/opencv/opencv_contrib/blob/4.x/modules/ximgproc/src/selectivesearchsegmentation.cpp

[4] V. A/S, “videometer-toolbox-python.” [Online]. Available: https://github.com/Videometer/videometer-toolbox-python

[5] Pytorch, “Vgg documentation.” [Online]. Available: https://pytorch.org/vision/main/models/vgg.html

[6] T. Bezdan and N. Bacanin, “Convolutional neural network layers and architectures,” 01 2019, pp. 445–451.