Introduction

Object detection and classification is a classic problem in computer vision. It has been widely applied in various fields, e.g., biometrics, surveillance, industry production control. Object detection usually requires output the bounding box of detected object and the label of its object type. On the other hand, image classification requires to produce object type(s) as the output for a given input image.

Objectives

In this project, I attempt to apply object detection and image classification techniques to detect oil palm seeds that are present in an image and classify each of them into category of good or bad quality.

Dataset

There are in total 110 images of good quality oil palm seeds and 105 images of bad quality seeds. Each image contains 10 oil palm seeds of the same category. The images are stored in two different folders "/GoodSeed" and "/BadSeed" which are accessible at here.

You should keep a separate testing dataset which must contain 20 images of good quality seeds and 20 images of bad quality seeds, respectively. The specific image files that should be included in the testing dataset are specified in "testdata.csv" file accessible at here.

Requirements

1. Design a method to locate each oil palm seed present in an image and classify it into the category of good or bad quality.
2. The method should be able to output a bounding box (rectangular or circle) for each of the oil palm seed detected.
3. The method should be able to output a label as to whether the oil palm seed detected is of good quality or bad quality. The label can either be text-based, or colour based.
4. Evaluate the method both quantitatively and qualitatively for detection and classification accuracy.
5. Design new method for locating and classifying oil palm seeds.
6. Only use the training dataset for designing and developing the method.
7. Use the testing dataset for evaluating the method. The ground truth of the testing dataset should not be part of the input to the program and should only be used as part of the evaluation on the results obtained for the testing dataset. .

Table of Contents

Introduction

The determination of good and bad palm oil seeds adopts several well-known computer vision techniques such as Histogram of Oriented Gradients (HOG), Local Binary Patterns (LBP), Scale Invariant Feature Transform (SIFT) and the usage of Support Vector Machines (SVM). The report is hence sectioned with several key features implemented, results analysis and a rigorous discussion on the implemented approaches.

Key Features Implemented

General

Some functions include display parameters to allow user to preview output images in Google Colab directly. We have also written our customised imshow function so that images are displayed much quicker compared to the original cv2_imshow function.

Seed Detection

In order to detect seeds, a binary seed mask is required. Our seed detection mainly follows the general object detection pipeline, which includes image pre-processing, segmentation and detection visualisation.

RGB images are first converted into HSV colour space. One key implementation of this pre-processing stage is the whole seed mask is obtained from three different parts and later combined. The first part is getting the general seed mask by doing Otsu's thresholding on the saturation channel of the HSV image. Seed edges are obtained by finding the difference between the value channel of the HSV image and the Gaussian-blurred dilated image. Canny edge detection is also performed to get the third part of the seed. For the final seed mask, morphological closing and holes filling are performed.

After having the seed mask, seed edges are calculated using contours. A bounding box is drawn for each detected seed and is used to crop them out as individual images.

As there is no labelled dataset for seed detection, the ground truth is prepared manually by annotating the bounding box for each seed in the test set using an online annotation tool (www.makesense.ai). The detection accuracy is then evaluated using intersection over union (IOU) metric. Finally, both ground truth and predicted bounding boxes are drawn to visualise the seed detection.

Seed Classification

It was found that the largest and smallest height x width for the seed images has a combined dimension of 548 x 859 and 200 x 187 respectively. To obtain a consistent result across all feature descriptors, each seed image is resized with the dimension 256 x 352. Additionally, the colour space for the images is converted from RGB to YCrBr as the latter processes the colour space into more meaningful information. The HOG implemented uses a 16 x 16 pixels per cell and a 10 x 10 cells per block. The LBP implemented uses a normalized histogram. Given the varying number of descriptors SIFT produces for each image, we have decided to generate a histogram based on its descriptors for every image to obtain a standardised set of descriptors for each image. This will also ensure its usability by the SVM as such was the requirement for SVM classifications.

The descriptors were stored in two datasets, a testing dataset containing the first 200 individual good seeds and the first 200 individual bad seeds, and the training dataset containing the remaining 1750 individual bad and good seeds. The training dataset was then randomized to minimize the chances of having a biased dataset and increasing the chances of having a representative dataset for training (Victor, 2019).

The 3 training dataset feature descriptors were fitted into the SVM classifier individually to obtain 3 different SVM classifiers. SVM was preferred as it requires significantly smaller dataset (Raschka, 2016) compared to convolutional neural networks. Finally, the classification results were obtained in two different ways. Firstly, the SVM classifiers were tested on their respective testing dataset and the accuracy for all 3 were obtained. Additionally, the mean of the individual seed classification results of the HoG and LBP feature descriptors were calculated and converted back to binary values to form the classification results for HoG+LBP features. The classification of each seed of the testing dataset was represented with an output bounding box labelled based on classification, where green indicates the seed is good and red indicates the seed is bad.

Results Analysis

Seed Detection (Quantitative + Qualitative)

There are 200 bad seeds and 201 good seeds in the test set. Out of these, 400 seeds are detected, which gives an accuracy of 99.75%. Using IOU metric of measuring the accuracy of seed area detection, the algorithm achieves an accuracy of 93.28%.

Figure 1 and 2 show examples of seed detection visualisation using bounding boxes, where green boxes are the ground truths and red boxes are the predictions of seed detection. All seeds in these two figures are detected but one is not in Figure 2. Most of the predicted bounding boxes are quite close to their ground truths, as shown in large overlapping area of those two boxes.

Figure 1: Seed detection visualisation for BadSeed19.jpg.

Figure 2: Seed detection visualisation for GoodSeed19.jpg.

Seed Classifications (Quantitative + Qualitative)

Method (s)	Number of Features Extracted	Accuracy Obtained (%)	Precision	Recall	F1-Score	SVM Compute Time (s)	Correctly Predicted Good Seeds	Correctly Predicted Bad Seeds
HOG	77760	76.3	0.76	0.76	0.76	671	160	145
LBP	122	72.8	0.73	0.73	0.73	0.616	150	141
SIFT	128	61.3	0.61	0.61	0.61	0.759	129	113
HOG + LBP + SIFT	-	76.3	0.76	0.76	0.76	-	157	149

Table 1: Classification and Feature Extraction results

Overall, HOG produces the highest accuracy in classifying the seeds with an accuracy of 76.2%. The hypothesis that a combination of HOG, LBP and SIFT descriptors would further enhance the classification accuracy of the SVM is proven true as shown in the table, albeit by a slim improvement. Furthermore, correctly predicted good seeds is on average higher than that of bad seeds, given that good seeds tend to have larger roots in width, length and height. However, in the dataset there are certain good seeds with smaller roots relative to other good seeds *Figure 5* and there are bad seeds with larger roots relative to other bad seeds *Figure 6*. Therefore, it is possible that the feature descriptors produced for these two cases are similar and hence misclassifying, bringing the accuracy results down. Figures 5 and 6 shows examples of seed classification visualisation using bounding boxes, where green boxes are the seeds classified as good and red boxes are the seeds classified as bad.

Figure 3: GoodSeed17_9.jpg

Figure 4: BadSeed1_0.jpg

Figure 5: Seed classification visualisation for GoodSeed0.jpg

Figure 6: Seed classification visualisation for BadSeed0.jpg

Additionally, we have tested several colour spaces of the images to be used in the HOG and it was found that YCrBr colour space performs the best, followed by grayscale images. On the other hand, HOG descriptors obtained from small pixels per cell tend to contain too much unnecessary information while too large of a pixels per cell will instead remove important features necessary for the SVM to classify as seen in figures 8 – 11. Lastly, figures 12 – 14 shows how radius affects the LBP descriptors.

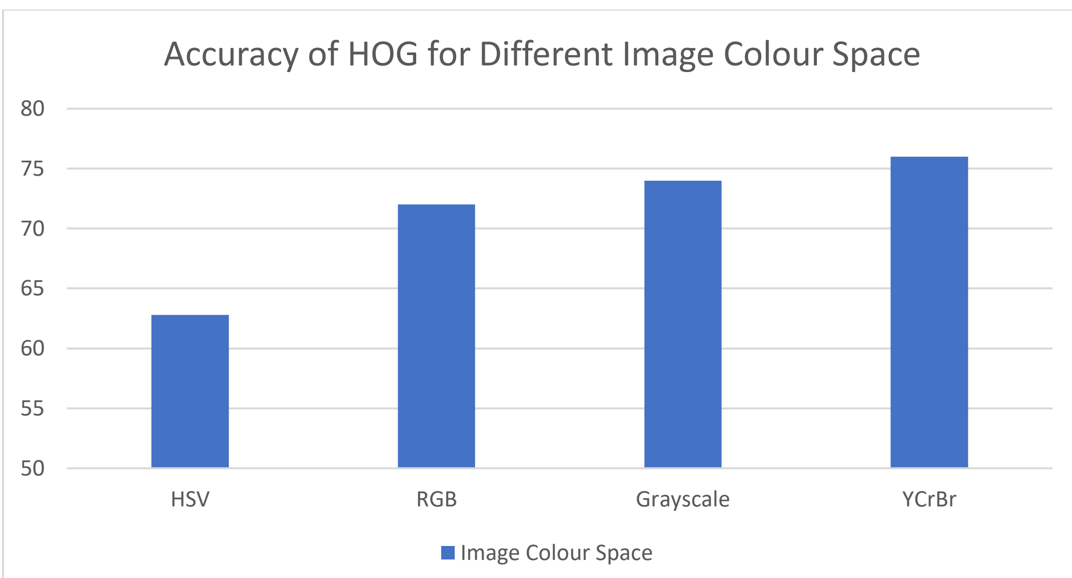

Figure 7: Accuracy of HoG for different Image Colour

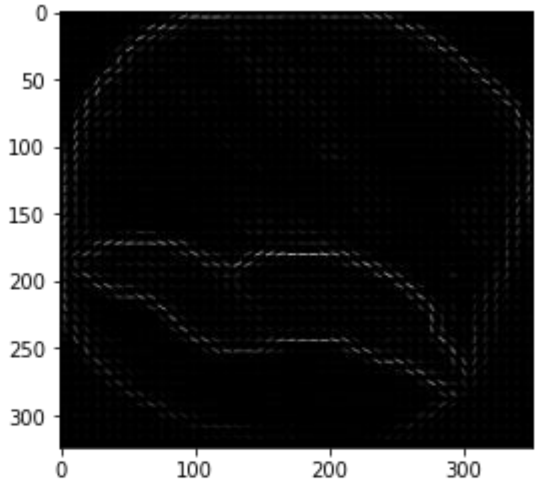

Figure 8: HOG with 8 x 8 pixels per cell

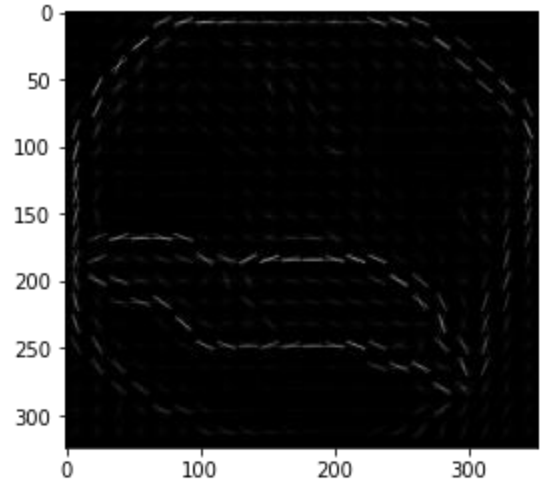

Figure 9: HOG with 16 x 16 pixels per cell

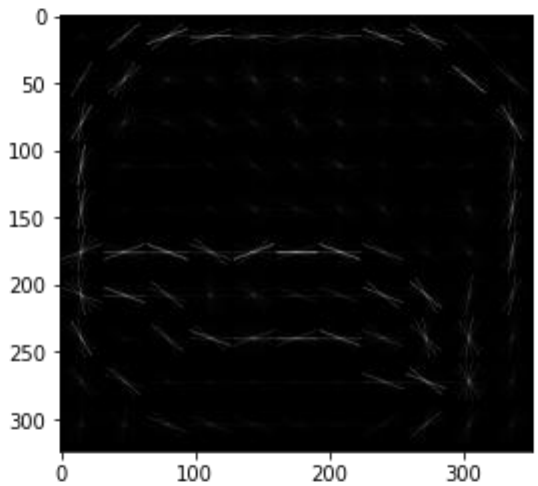

Figure 10: HOG with 32 x 32 pixels per cell

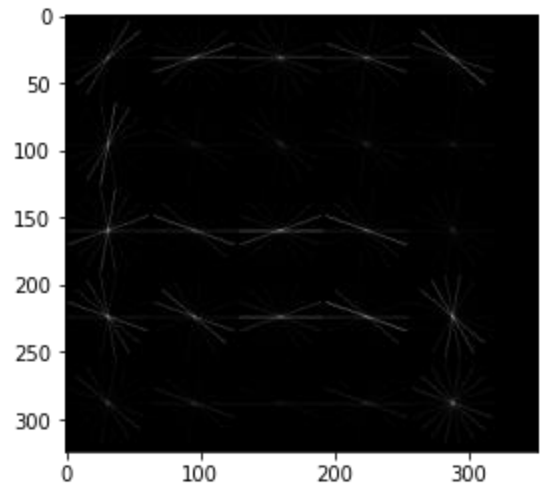

Figure 11: HOG with 64 x 64 pixels per cell

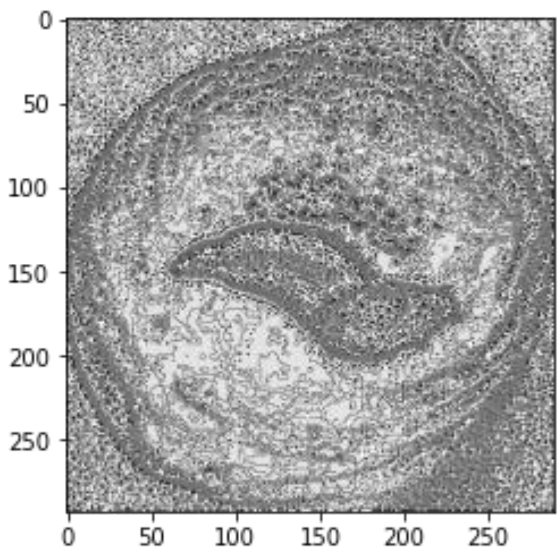

Figure 12: LBP with radius = 1

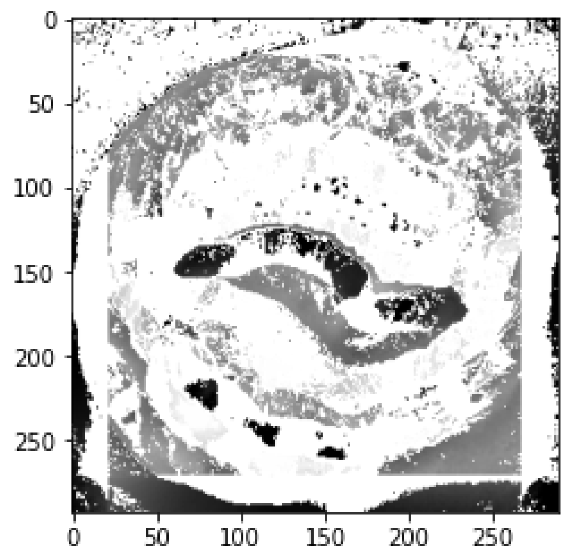

Figure 13: LBP with radius = 22

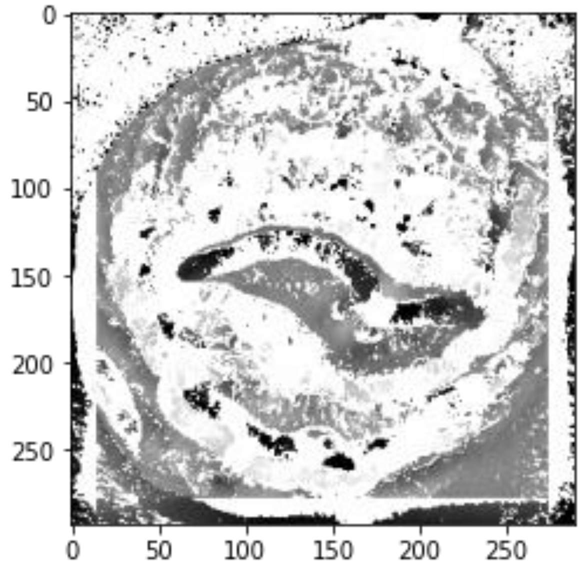

Figure 14: LBP with radius = 22

Discussions

HSV colour space is used during the pre-processing stage as it separates luminance from colour information compared to RGB colour space (Sural, et al., 2002) & (Bora, et al., 2008). Using HSV gives a much better starting point for seed separation from the background, which can be clearly seen in the saturation channel of the HSV image. This is also why Otsu's thresholding is performed on this channel for the first part of seed mask generation.

The rationale behind combining three different parts to form the whole seed mask is because each part only obtains specific parts of the seed. It becomes obvious that combining all three parts will give a more complete seed mask by taking the strengths of each part's method. The second and third part of the seed mask are obtaining the seed edges. However, the difference between these two parts is the second part gives very clear edges due to the difference computation while Canny edge detector uses two thresholds to detect strong and weak edges, which gives some parts of the seed body in addition to the edges.

Morphological closing is done a few times throughout the pre-processing stage with a rectangular kernel even though the seed is generally round. This is because we aim to have bounding boxes in the end and not the exact shape of the seed for seed detection. And using rectangular kernel is much faster than eclipse kernel due to the decomposition of the rectangle to have horizontal and vertical segments, which makes computations to be able to run simultaneously (Dokladal & Dokladalova, 2011).

IOU is used as the seed detection metric due to its simplicity for implementation and it is the most popular object detection benchmarks metric (Rezatofighi, et al., 2019). Nonetheless, IOU is theoretically proven to be bias towards large scale objects in the dataset (Hager, et al., 2018). It is good to note this bias for other object detections but this is not applicable in our task as the seeds are all almost the similar size.

Our seed detection algorithm manages to achieve an accuracy of 93.28%. Some of the errors may be caused by the presence of shadow and lighting where the algorithm mistakenly assumes the shadow as part of the seed, given the similar shade of colour. It should be noted however, that it was also a challenge for the human eye to properly determine the shadows from the seeds when attempting manual labelling of the seeds.

One limitation for our seed detection algorithm is that we can only detection a maximum of 10 seeds in a single image, as stated in the coursework issue sheet that there are only 10 seeds in each image. This is because we take the largest 10 seed contours to remove any noise when the bounding boxes are generated. However, it would be something we can work on in the future to extend the number of possible detected seeds to any number.

Seed resizing is an integral part of the feature extraction as it is computationally too expensive to use a large seed image. Yet, a smaller seed image proves to be ineffective, given the lack of useful features. Hence, a balance of performance and computational cost were determined by using half of the largest height and one third of the largest width.

Although it was suggested by (Banerji, et al., 2013) that using YCrBr colour space for HOG descriptors will perform significantly better than RGB or grayscale, this was not the case as reported in *figure 7*. Instead, they have suggested that grayscale images tend to perform worse for HOG descriptors compared to HSV. Our results have shown otherwise, and this may be caused by the fact that the seed has 2 dominant colours. The black shade of the seed's husk and the white shades of the seed's roots. Furthermore, we have generated a few black images upon detection of an invalid seed image to be used as noise. This will further help reduce overfitting in the SVM. Given that the Y dimension in YCrBr represents the grayscale image, it is therefore assumed that the contribution of grayscale dimension shows positive results.

The current approach of combining each descriptor to determine the quality of the seeds may suffer from a 'majority rule' problem where the classified candidate may not be necessary 'good' even though it won by majority. This may be especially prevalent given that HOG outperforms LBP and SIFT on an individual scale, but it is likely that the worse performing LBP and SIFT misclassifies seed images, resulting in an overall misclassified seed. To overcome such issue, Borda count approach may be introduced when determining seeds. That is, the seed classification by each descriptor is given a weightage where HOG is given priority when determining the classification over LBP and SIFT. Hence, it would possibly require all 3 descriptors unanimously determine a seed's classification. Upon which, giving HOG a higher priority.

The front (sprouts) of the seed varies in position and angles. LBP runs well on images where the object of interest is mainly posing front, and decreases in effectiveness as the dataset gets complicated and the algorithm cannot extract the features completely (Raj & Niar, 2017). This may be the reason for LBP's low accuracy result, as seen in Figure 12 and 13 where the root are at the front vs top respectively. One possible improvement is to adopt a late fusion (Trong, et al., 2020) approach where the roots and seeds are further segmented to produce another set of feature descriptor upon which a combination is used in the SVM (Rainville, et al., 2012).

Figure 12: GoodSeed10_4.jpg

Figure 13: GoodSeed9_6.jpg

HOG descriptors are variant against angle of rotation (Fleyeh & Roch, 2013), therefore disoriented seeds might not have been classified correctly. As seen in Figure 7 and Figure 8, although both seeds are of the same class, good, their orientation is different. A possible solution to identify the orientation of the root and segment the seeds into their respective orientations before performing feature extraction and classification.

Although LBP can produce a relatively high accuracy, much of its weaknesses lies in its sensitivity to noise and its radius parameter (Parveen, et al., 2016). That is, a smaller radius tends to miss out the larger areas of the seed husk where most of its region has the same shade of gray whereas an extremely large radius will use the roots' threshold value when computing the surrounding husk area. While determining a good radius proves to be a challenge and remains to be an optimization problem to be solved in the future, we have successfully determined the current radius of size 15 to be the best via brute force. Hence, it is without a doubt that the tuning of LBP requires extensive work and time.

Several issues arise with the usage of SIFT descriptors. Firstly, the nature of the descriptors where they are mainly used for object detection and object matching. As each of the seed images were not matched to another similar image, the results of SIFT was lesser than hypothesized. Secondly, SIFT produces varying numbers of descriptors for each individual image where an SVM cannot classify such inputs. Hence, the histogram built on each image's descriptors were reduced significantly. An alternative for this is Dense SIFT, but the issue with this algorithm is that it computes a huge number of descriptors, which may not be computationally feasible given an image size of 256 x 352 x 128 descriptors, having a total of 11534336 descriptors for a single seed image. Although it may be possible to resize to a smaller image, the accuracy obtained is still extremely low. It is therefore not surprising to obtain a weaker performance when using SIFT in classifying the seeds.

The main caveat of SVM is the high computational complexity, which is mainly influenced by the dimension of feature vector input (Lee, et al., 2015). However, as the dimension of feature vector increases, we can expect better accuracy and higher computational time (Lee, et al., 2015). This explains why HOG (77760 features) performed 21.8% higher than LBP (18 features) and SIFT (128 features), and had a higher classification computational time (656 seconds) compared to LBP (0.177 seconds) and SIFT (0.759 seconds)

Aggregating the individual class predictions of an even number of classifiers by computing the mean brings up a weakness where a bias exists towards having equally predicted good and bad seeds classified as good seeds, as rounding 0.5 to the nearest integral value results in the seeds classified as good. This problem does not persist in our approach as we used an odd number of classifiers used, in this case 3.

Conclusion

The proposed methods for classifications have been successfully implemented, although with many questions still left unanswered with many possible future works in the open. One such question would be the performance of the implemented descriptors on different palm seed species and varying lighting conditions. Different classifiers such as convolutional neural network could also be applied when a larger dataset is obtained as the methods proposed are framed for a small dataset and the reductionistic assumption that only seeds are present in each image.

Appendix

Appendix A

Feature Extraction Computational Time for SIFT, LBP and HoG.

```
sift_output = feature_ext('SIFT')

SIFT feature extraction: 100% 214/214 [01:27<00:00, 2.45image/s]
seed/GoodSeedCropped/GoodSeed37_1.jpg is not a valid seed image!
seed/GoodSeedCropped/GoodSeed62_7.jpg is not a valid seed image!
seed/GoodSeedCropped/GoodSeed88_3.jpg is not a valid seed image!
seed/GoodSeedCropped/GoodSeed89_3.jpg is not a valid seed image!
seed/GoodSeedCropped/GoodSeed65_9.jpg is not a valid seed image!

time: 1min 27s (started: 2021-04-09 02:08:17 +08:00)

[20] lbp_output = feature_ext('LBP')

LBP feature extraction: 100% 214/214 [01:35<00:00, 2.24image/s]
seed/GoodSeedCropped/GoodSeed37_1.jpg is not a valid seed image!
seed/GoodSeedCropped/GoodSeed62_7.jpg is not a valid seed image!
seed/GoodSeedCropped/GoodSeed88_3.jpg is not a valid seed image!
seed/GoodSeedCropped/GoodSeed89_3.jpg is not a valid seed image!
seed/GoodSeedCropped/GoodSeed65_9.jpg is not a valid seed image!

time: 1min 35s (started: 2021-04-09 02:10:32 +08:00)

# Call the feature extraction function. Input the argument as 'HoG' or 'LBP' or '
hog_output = feature_ext('HoG')
#lbp_output = feature_ext('LBP')
#sift_output = feature_ext('SIFT')

HoG feature extraction: 100% 214/214 [01:38<00:00, 2.18image/s]
seed/GoodSeedCropped/GoodSeed37_1.jpg is not a valid seed image!
seed/GoodSeedCropped/GoodSeed62_7.jpg is not a valid seed image!
seed/GoodSeedCropped/GoodSeed88_3.jpg is not a valid seed image!
seed/GoodSeedCropped/GoodSeed89_3.jpg is not a valid seed image!
seed/GoodSeedCropped/GoodSeed65_9.jpg is not a valid seed image!

time: 1min 38s (started: 2021-04-09 02:12:24 +08:00)
```

Appendix B

SVM Classification Computational Time for HoG, LBP and SIFT.

```
_, hog_y_pred = svm_classifier(data['hog']['train']['descriptor'], data['hog']['train']['label'], data['hog']['test']['descriptor'])

time: 11min 2s (started: 2021-04-09 02:17:37 +08:00)

_, lbp_y_pred = svm_classifier(data['lbp']['train']['descriptor'], data['lbp']['train']['label'], data['lbp']['test']['descriptor'])

time: 188 ms (started: 2021-04-09 02:17:29 +08:00)

[25] _, sift_y_pred = svm_classifier(data['sift']['train']['descriptor'], data['sift']['train']['label'], data['sift']['test']['descriptor'])

time: 759 ms (started: 2021-04-09 02:17:14 +08:00)
```

Appendix C

SVM Classification using HOG, LBP, SIFT and Combined Features.

```
==================================================

Classification using HOG

Accuracy: 0.7625

               precision    recall  f1-score   support

 Bad Seed (0)       0.78      0.72      0.75       200
Good Seed (1)       0.74      0.80      0.77       200

     accuracy                           0.76       400
    macro avg       0.76      0.76      0.76       400
 weighted avg       0.76      0.76      0.76       400

==================================================

Classification using LBP

Accuracy: 0.5475

               precision    recall  f1-score   support

 Bad Seed (0)       0.56      0.45      0.50       200
Good Seed (1)       0.54      0.65      0.59       200

     accuracy                           0.55       400
    macro avg       0.55      0.55      0.54       400
 weighted avg       0.55      0.55      0.54       400

==================================================

Classification using SIFT

Accuracy: 0.6125

               precision    recall  f1-score   support

 Bad Seed (0)       0.62      0.56      0.59       200
Good Seed (1)       0.60      0.66      0.63       200

     accuracy                           0.61       400
    macro avg       0.61      0.61      0.61       400
 weighted avg       0.61      0.61      0.61       400

==================================================

Classification using HOG, LBP and SIFT

Accuracy: 0.655

               precision    recall  f1-score   support

 Bad Seed (0)       0.69      0.56      0.62       200
Good Seed (1)       0.63      0.74      0.68       200

     accuracy                           0.66       400
    macro avg       0.66      0.66      0.65       400
 weighted avg       0.66      0.66      0.65       400
```

Appendix D

Classification results for HOG with different colour space

Classification using HOG

Accuracy: 0.74

	precision	recall	f1-score	support
Bad Seed (0)	0.76	0.70	0.73	200
Good Seed (1)	0.72	0.78	0.75	200
accuracy			0.74	400
macro avg	0.74	0.74	0.74	400
weighted avg	0.74	0.74	0.74	400

Figure 1: HOG Classification results using Grayscale images

Classification using HOG

Accuracy: 0.76

	precision	recall	f1-score	support
Bad Seed (0)	0.79	0.71	0.75	200
Good Seed (1)	0.74	0.81	0.77	200
accuracy			0.76	400
macro avg	0.76	0.76	0.76	400
weighted avg	0.76	0.76	0.76	400

Figure 2: HOG Classification results using YCrBr images

Classification using HOG

Accuracy: 0.6275

	precision	recall	f1-score	support
Bad Seed (0)	0.64	0.58	0.61	200
Good Seed (1)	0.62	0.67	0.64	200
accuracy			0.63	400
macro avg	0.63	0.63	0.63	400
weighted avg	0.63	0.63	0.63	400

Figure 3: HOG Classification results using HSV images

References

Banerji, S., Sinha, A. & Liu, C., 2013. *HaarHOG: Improving the HOG Descriptor for Image Classification.* Manchester, IEEE, pp. 4276-4281.

Bora, D. J., Gupta, A. K. & Khan, F. A., 2008. Comparing the Performance of L*A*B* and HSV Color Spaces with Respect to Color Image Segmentation. *International Journal of Emerging Technology and Advanced Engineering,* 5(2), pp. 192-203.

Dokladal, P. & Dokladalova, E., 2011. Computationally efficient, one-pass algorithm for morphological filters. *Journal of Visual Communication and Image Representation,* 22(5), pp. 411-420.

Fleyeh, H. & Roch, J., 2013. BENCHMARK EVALUATION OF HOG DESCRIPTORS AS FEATURES FOR CLASSIFICATION OF TRAFFIC SIGNS. *International Journal for Traffic and Transport Engineering,* pp. 448-464.

Hager, G., Felsberg, M. & Khan, F. S., 2018. *Countering bias in tracking evaluations.* s.l., Science and Technology Publications, Lda.

Lee, S.-H., Bang, M., Jung, K.-H. & Yi, K., 2015. An efficient selection of HOG feature for SVM classification of vehicle. *2015 International Symposium on Consumer Electronics (ISCE),* pp. 1-2.

Parveen, S., Naeem, N. & Devi, J., 2016. REVIEW ON LOCAL BINARY PATTERN (LBP) TEXTURE DESCRIPTOR AND ITS VARIANTS.. *International Journal of Advanced Research (IJAR),* pp. 708-717.

Rainville, F.-M.D. et al., 2012. Bayesian classification and unsupervised learning for isolating weeds in row crops. *Industrial and Commercial Application,* pp. 401-414.

Raj, S. N. & Niar, P., 2017. Comparison Study of Algorithms Used for Feature Extraction in Facial Recognition.. *International Journal of Computer Science and Information Technologies,* 8(2), pp. 163-166.

Rezatofighi, H. et al., 2019. *Generalized intersection over union: A metric and a loss for bounding box regression.* s.l., s.n.

Sural, S., Qian, G. & Pramanik, S., 2002. *Segmentation and histogram generation using the HSV color space for image retrieval.* s.l., IEEE, pp. II--II.

Trong, V. H., Gwang-hyun, Y., Vu, D. T. & Jin-young, K., 2020. Late fusion of multimodal deep neural networks for weeds classification. *Computers and Electronics in Agriculture,* pp. 1-13.