EXECUTIVE SUMMARY OF THE THESIS

Blob analysis for dimensional features and shape defect extraction in agri-food production

TESI MAGISTRALE IN FOOD ENGINEERING – INGEGNERIA ALIMENTARE

AUTHOR: DAVIDE MONDIN

ADVISOR: MARCO TARABINI

ACADEMIC YEAR: 2022-2023

1. Introduction

This Master thesis deal with BLOB analysis, a technique used to classify extracted objects leveraging on their properties (like colour, shape and size). The thesis' aim is to locate aubergine slices from RBG images obtained from a conveyor belt, describe the extracted features, and compare each other. A Deep learning technique named Yolov8 and a machine learning application named Random Forest have been applied in order to extract the targets. The main challenges encountered throughout the whole process were: overlapped objects isolation, distinguish pixels belonging to aubergine skin when compared to background and artificial light reflects. The aim of this work is to provide an early stages workflow for those food manufacturers interested in applying remote sensing technologies based on deep learning and testing RGB colour channels for their potential inclusion in multispectral imaging applications within the TESORO (Studio e sviluppo di TEcnologie avanzate per il SORting automaticO nei processi di produzione industriali) project.

1.1 State of the art

Yolo (you only look once) is a deep learning technology used in the field of machine vision for object recognition and classification. Since its first appearance [1], Yolo and followings have been state-of-the art online applications due to their rapid inference times and good precision. Yolov8 is one of the most recent versions, it has been released by *Ultralytics* in latter half of 2022. Yolov8x is the subvariant used in this work due to its higher Precision/Recall performances with respect the other subvariants.

Few ML algorithms were feed with pixel values represented in 4 different colourspaces (RGB, HSV, Luv and YCbCr), building up a 12 features model, the tested techniques are: SGD, Decision Tree, Random Forest, Softmax Regression, K-means and DBscan [2].

2. Methods

The workflow followed in this work is made of three parts, the first involves Yolov8 as well as the necessary procedures to setup the dataset, the second is dedicated to set up and run ML algorithms with the purpose of sorting pixels in aubergines classes or background, the output of the ML algorithm that reveals better performances is combined with the results of Yolov8 in order to filter out background pixels, the third step is focused on features extraction applied to the BLOBs obtained in the previous phase as well as the description and comparison of such features, the whole workflow is summarized in Figure 2.1.

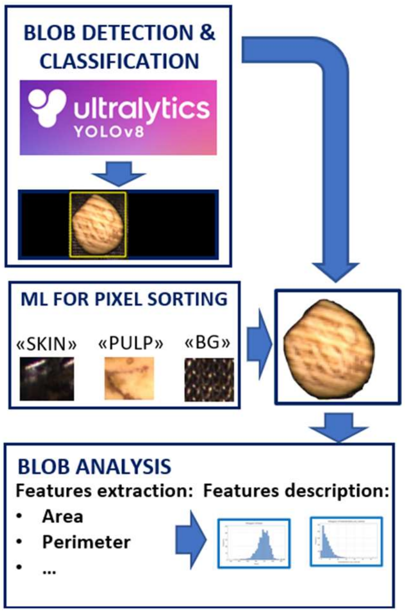

Figure 2.1: The workflow.

The images used in this thesis has been acquired in a real food production plant called *O.P. COTRAPA 2000*, an Italian food manufacturer and partner of the TESORO project collaborating with *Politecnico di Milano*.

Aubergine slices were acquired under two conditions, the first condition ("compliant" from now) refers to slices acquired downstream of the sorting process carried out by human operators, here the product is almost completely compliant and properly distributed on the conveyor belt in term of quantity and without significantly overlapped slices, 824 images of this type are available, one of them is show in Figure 2.2. The second part of the dataset was acquired under a different condition (named "discard" from now), it is made of 256 pictures obtained arresting the sorting process and loading an high quantity of product, such product was classified as non-compliant by the operators working on that specific line, as a consequence, the majority of the product doesn't satisfy the quality requirements chosen by the company, an example is shown in Figure 2.3. None of the images used in this work represents the expected "to be" working condition, which is similar to "Compliant" dataset in term of product arrangement but the relative frequency of the product to be discarded would be higher.

Figure 2.2: "Compliant" image sample.

Figure 2.3: "Discard" image sample.

The classes established for this work are the following:

- Back/Top: aubergine slices exposing the skin to the camera.
- Speckled: those products having one or more significant black spot/s.
- Broken/Holes: this class includes slices that are unintentionally broken or presenting other kind of discontinuities, like holes.
- Other/Compliant: all the other objects released on the conveyor belt, including the "good" products.

2.1 Yolo

The Yolov8's training/validation/test datasets has been labelled using *LabelImg*, it's a software that gives the possibility to save the annotation files in "YOLO" format, it is the standard used by all Yolo releases. The dataset has been split following the usual proportion:

- 70% for "training set".
- 20% for "validation set".
- 10% for "test set".

The "training set" and "validation set" are made by 50/50 mix of "discard" and "compliant" images while "test set" is made solely by "complaint" images. Although the inference on "discard" images lead to poor results due to the high density of overlapped products, they have been included in the "training set" and "validation set" because of their preciousness in shortage conditions. The material used to train and test Yolov8x is summarized in Table 2.1.

Table 2.1: Yolov8 dataset composition.

Set split	Compliant	Discard	Total
Train	179	179	358
Validation	51	51	102
Test	51	0	51

The obtained neural network is used as a prediction tool for the entire "compliant" dataset, the output can be conveniently store as a list of Yolo annotation files, one for each processed image.

2.2 Pixel sorting

Some machine learning algorithms were trained to predict whether a pixel belongs to the product or not, in order to facilitate the candidate models and understand the results, two classes have been assigned to the foreground: aubergine pixels are sorted in "Pulp" and "Skin" classes, they represent the internal part of the product and the vegetable skin respectively. Most of the proposed machine learning techniques are supervised, therefore, they require a ground truth to be trained. 50 pictures from "compliant" dataset were selected and two masks for each of them were drawn, these masks cover the "Pulp" and "Skin" pixels. The pixel values are shuffled and split in "train set" and "test set" following a 90/10 proportion, the machine learning tools have been fed with the pixel values, the most promising technique has been selected after the performance comparison, F score has been used as deciding performance for the supervised models.

The best ML algorithm is used to filter out the useless pixels form the rectangle located by Yolov8x but isn't enough to isolate the target, some tools has been applied in order to properly isolate only the pixels belonging to the target object.

2.3 Features extraction

The extraction process starts when Yolov8x locates the aubergine slice and its class, for the way Yolov8x works, it identifies the object through a rectangle, from this rectangle, the pixels are filtered according with the predicted class. The products belonging to "Back/Top" is made almost exclusively by "Skin" pixels whereas the other classes are associated to "Pulp" pixels, hence only one type of pixel is extracted from the drawn rectangle. This procedure is accomplished for every Yolov8x's output line unless the box's centre fall within 250 pixels from the image's vertical border or 75 pixels from the horizontal border, those products are excluded because is challenging to reconstruct the true shape for those products that aren't completely included in the picture. The BLOB isn't extracted immediately since few problems still in place, in fact, Yolov8x' rectangle may include more than one complete slice or exclude parts of the object, moreover, "Back/Top" products are affected by a relevant lack of pixels due to presence of reflects.

Only the largest BLOB inside the rectangle is chosen and "expanded", this means that the pixels contiguous to the BLOB and belonging to the same pixel class are included in the BLOB itself, such procedure tackle those cases where the aubergine slice is partially located outside the rectangle; this solution introduces an extra problem if the target is overlapped (or extremely close) with one or more slices of the same pixel class, along with the target, the neighbouring slice/slices may be fully enclosed in the BLOB. The previously obtained BLOB undergoes to "closing".

Watershed transformation is applied in order to separate the target from the neighbouring products, the original watershed transformation was modified in the nuclei extraction method to overcome the unperfect circle-like shape of the slices, the variation adopted in this case starts after the distance map computation, the distance map has been processed by means of a median filter and the local maxima of such map are recorded, the nuclei are originated from circles located in the local maxima having a diameter equal to a half of the pick value. Watershed transformation generates multiple BLOBs, only one of them is the target, the BLOB being closer to the Yolov8x rectangle centre is chosen.

Once target BLOB is located, dimensional and shape-based features can be measured, all the residual features are extracted superimposing the BLOB to the original RGB picture. A list of the extracted features is provided:

- **Box Distance**: Is the centre-centre distance between the rectangle circumscribed to the BLOB and the rectangle coming from Yolov8x; the unit of measure is "pixels".
- **Area**: Number of pixels that makes the BLOB.
- **Perimeter**: BLOB perimeter length in pixels.
- **Circularity**: It measures how the BLOB is similar to a circle.
- **Solidity**: It measures the BLOB convexity. Is the ratio between **Area** and Hull area.
- **Extent**: Is the ratio between **Area** and the area of the rectangle circumscribed to the BLOB.
- **G_avg**: Is the average value in the green channel.
- **G_stddev**: Is the sample standard deviation in the green channel.
- **B_avg**: Is the average value in the blue channel.
- **B_stddev**: Is the sample standard deviation in the blue channel.
- **R_avg**: Is the average value in the red channel.
- **R_stddev**: Is the sample standard deviation in the red channel.
- **H_avg**: Is the average value in the hue channel.
- **H_stddev**: Is the sample standard deviation in the hue channel.
- **S_avg**: Is the average value in the saturation channel.
- **S_stddev**: Is the sample standard deviation in the saturation channel.
- **V_avg**: Is the average value in the V channel.
- **V_stddev**: Is the sample standard deviation in the V channel.
- **Major_axis**: Is the major axis length of the ellipse that approximates the BLOB, measured in pixels. The elliptical approximation is carried out by *opencv.fitEllipse* function.
- **Minor_axis**: Is the minor axis length of the ellipse that approximates the BLOB, measured in pixels. The elliptical approximation is carried out by *opencv.fitEllipse* function.
- **Inertia_X**: Is the second momentum of the BLOB, calculated in BLOB's centre with respect to the X axis passing through the centroid.
- **Inertia_Y**: Is the second momentum of the BLOB, calculated in BLOB's centre with respect to the X axis passing through the centroid.

The features are collected and sorted by class. For each feature-class combination the sample mean and the sample standard deviation are calculated, such data are useful to describe the products, moreover the features are potentially able to distinguish between the classes. In this thesis, the t-test is applied to each class couple in order to test whether or not their mean is the same, and this is done for each feature. All the statistical tests are played with alpha equal to 5%.

3. Results

3.1 Yolo results

Yolov8x training process has been stopped after 81 training epochs, the resulting cross-class confusion matrix is shown in Figure 3.1.

Conf. matrix		True				
		Back/Top	Speckled	Holes/Bro...	Other/Com...	BG
Predicted	Back/Top	0.9	0.00	0.00	0.01	0.30
	Speckled	0.01	0.49	0.24	0.02	0.16
	Holes/Bro...	0.01	0.01	0.21	0.01	0.07
	Other/Com..	0.03	0.47	0.56	0.92	0.48
	BG	0.06	0.03	0.00	0.03	-

Figure 3.1:Yolov8x confusion matrix, values as percentage of the true instances for that class.

Yolov8x was able to detect "Back/Top" instances, it is expected since this class is well represented in the dataset and is fairly different with respect to the other classes, 10 % of the instances belonging to "Back/Top" are misclassified and, in those cases, half of the times are classified as “Background”. “Speckled” and “Holes/Broken” classes are characterized by poor performances, “Speckled” instances are misclassified as “Other/Compliant” half of the times while “Holes/Broken” are predicted as “Other/Compliant” in 56% of the cases, the remaining are almost evenly distributed between the true positive (“Holes/Broken”) and “Speckled” class; “compliant” products are retrieved 92% of the times. The localization performance is one of the best achievements in this case, the “box loss”, a performance measuring the deviation between the predicted box and the ground truth, is equal to 0.55.

3.2 Pixel sorting results

Random Forest performed better than the others ML techniques, the resulting F score is equal to 0.95, Random Forest reveled which are the features useless for the classification, for this reason, Random Forest has been played again with only the most 5 significant colour channels (R, u, v, Cb and Cr) without a significant loss of performances, Figure 3.2 shows the new performances by means of a confusion matrix and its most important performances.

RANDOM FOREST (5 FEATURES)

Confusion matrix		True		
		BG	Pulp	Skin
Predicted	BG	0.92	0.02	0.08
	Pulp	0.05	0.98	0.00
	Skin	0.03	0.00	0.92

	Precision	Recall	F-score
BG	0.92	0.97	0.94
Pulp	0.98	0.95	0.97
Skin	0.92	0.79	0.85
Weighted avg	0.95	0.95	0.95

Figure 3.2:Random Forest confusion matrix and performances, 5 features run.

The performance loss caused by the model’s dimensional reduction still limited if compared with the dimensional reduction degree, in fact, F score still equal to 0.95 and the recall is close the full dimension model’s one.

3.3 Features extraction results

The t-tests on feature means have been performed for each class couple, 22 features have been tested and the null hypothesis has been accepted 21 times over the 132 tests.

All but two of the t-tests involving “Back/Top” class have been rejected, it suggests this class as the most different with respected to the others and almost all the features may be used to characterize the class, the RGB colour averages still most characterizing ones due to their significant difference with respect to the other classes and their understandable meaning. The less valuable features are the ones related to the standard deviation in both the colourspaces; it worths for all the classes but “Back/Top”, as already mentioned, its characteristics aren’t equal to the relative counterparts, moreover, 11 over the 21 tests resulting in an acceptance of the null hypothesis have been caused by t-tests involving the standard deviation features.

4. Discussion and Conclusion

This thesis faces a deep learning application case that could be found when trying to adopt machine learning technologies in an ongoing manufacturing process. The proposed method sees the combination of deep learning and classical machine learning techniques to locate and classify defective products in aubergine production lines. As DL model, Yolov8 has been selected as it represents the state of the art for object detection purposes. Among the selected classed to be identified within the acquired dataset, “Speckled” and “Holes/Broken” classes are underrepresented in the “compliant” dataset, they account approximately 7.9% and 1.3% respectively. Yolov8x managed to retrieve more than 90% of the objects belonging to “Back/Top” and “Other/Compliant” classes while it struggles on distinguish the remaining classes.

The drawn boxes enclose properly the objects while the completely missed targets are only those slices located on picture’s borders, those products are barely visible. Yolov8x performances may be enhanced increasing the training pictures quantity, this operation still time consuming and introduces inconsistencies when carried out by multiple operators; adding or changing the channels may introduce additional useful information, in fact, RGB imaging looks effective in recognizing “Back/Top” aubergine slices while specific wavelength bands are proved to be in detecting rotten organic matter and fungal contaminations [3].

The second part regarding the application of ML techniques sees the comparison of different algorithms to classify images at pixel level in effective parts and background; among the ML techniques tested in this work, Random Forest has been selected to be used in the main workflow due to its higher performances when compared to other ML techniques. Random Forest’s performances are explained by the similarity of the background pixels and reflects on the product, although “Skin” pixels are severely affected by a low Recall value (Figure 3.2), they account for less than 7% (in “compliant” images), as a result, its impact on the weighted performances is low. This problem may be tackled adopting two approaches, the first requires changing the process such a way that the product’s skin is distinguishable to the background, replacing the conveyor or the lighting system itself. The second approach is based on wavelength selection, in fact, conveyor belt reflectivity may be significantly different to the organic matter’s one when acquiring in specific wavelengths [4].

Feature values have been successfully extracted; therefore, mean and standard deviation have been calculated, the test results show as the features have good potential when trying to guess the classes. The errors generated in the previous phases impact on t-test reliability since the afore mentioned tests are partially fed with highly deformed BLOBs, as shown in Figure 3.1, a relevant fraction of “Speckled” and “Holes/Broken” are misclassified, hence, a significant quantity of feature values are allocated to the wrong class.

The workflow proposed in this document (Figure 2.1) may be applied as set up procedure in industrial applications, although the limited effort put in labelling, the model is effective in locating the food products while the poor classification performances can be adjusted in two possible non-exclusive ways:

- Leveraging on operators’ knowledge on the features; some features are easy to understand from a human perspective, thus the operators could establish a few thresholds on features’ values with the purpose to assign the instance in the classes, this process operates on the workflow upstream fixing the incorrect class predictions that can be used to refeed a training process.
- A possible improvement is the addition of further image material to the already existing one, since a limiting resource is the labelling time, the existing model could be used to speed up this activity because the localisation is correct, and the operator is only responsible for selecting the class.

Future researches may investigate the usage of image channels different to the ones used in this work, starting from those wavelengths that showed good potential for food quality inspection, also for detecting quality criticalities different to what seen in in this case, RGB channels still useful to classify some objects (like “Back/Top” class), furthermore, the related sensors/filters are easy to access due to the industrial development experienced in RGB imaging.

References

[1] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *In Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016.

[2] A. Géron, "Hands-On machine learning with scikit-learn, keras & tensorflow," O'Reilly, 2019, pp. 114-124;149-151;177-184;199-200;240-249;256-259.

[3] J. Gòmez-Sanchìs, G. Camps-Valls, E. Moltò, L. Gòmez-Chova, N. Aleixos and J. Blasco, "Segmentation of hyperspectral images for the detection of rotten mandarins.," in *Analysis and Recognition: 5th International Conference*, Póvoa de Varzim, 2008.

[4] E. Minieri and E. Milani, "POLITesi," 4 Maggio 2023. [Online]. Available: https://hdl.handle.net/10589/212282. [Accessed 7 Novembre 2023].