Food Control 158 (2024) 110198

Contents lists available at ScienceDirect

Food Control

journal homepage: www.elsevier.com/locate/foodcont

Neural networks allow the automatic verification of the type of flour, analysing the starch granule morphology, to ensure the protected geographical indication 'Galician Bread'

Xosé R. Fdez-Vidal [a,*], Nerea Fernández-Canto [b], María Ángeles Romero-Rodríguez [c,e], Ana María Ramos-Cabrer [c,e], Santiago Pereira-Lorenzo [d,e], Matilde Lombardero-Fernández [b,e]

[a] Centro de Investigación en Tecnoloxías da Información (CITIUS), University of Santiago de Compostela, Campus Vida s/n, Santiago de Compostela, 15782, Galicia, Spain
[b] Unit of Veterinary Anatomy and Histology, Department of Anatomy, Animal Production and Clinical Veterinary Sciences, Faculty of Veterinary Sciences, University of Santiago de Compostela, Campus Terra, Lugo, 27002, Galicia, Spain
[c] Department of Analytical Chemistry, Nutrition and Bromatology, Faculty of Science, University of Santiago de Compostela, Campus Terra, Lugo, 27002, Galicia, Spain
[d] Department of Crop Production and Projects of Engineering, Higher Polytechnic School of Engineering, University of Santiago de Compostela, Campus Terra, Lugo, 27002, Galicia, Spain
[e] IBADER, University of Santiago de Compostela, Campus Terra, Lugo, 27002, Galicia, Spain

ARTICLE INFO

Keywords:
Flour starch granules
Polarized microscopy
Manual starch identification and counting
Automatic starch identification and counting
Neural networks
Elliptical fit
Instance segmentation
Mask R–CNN

ABSTRACT

Quality control of flour is essential to control the quality of bread produced from it. We propose a control method based on the morphological characteristics of the granules of starch. The automation of the identification, segmentation and determination of the average size of the granules of starch of each of the cereals that make up a flour, from microscopy images, is an essential procedure for producers who want to produce bread under the protected geographical indication (PGI) 'Galician Bread'. This identification and counting procedure, if performed manually, is a tedious activity for a trained expert, and is very time consuming. Thus, automating this task would streamline the process, in addition to saving a great deal of time. This paper addresses this problem by using deep learning approaches (Mask R–CNN) to predict the type of the granule of starch and its size for the first time. The trained models are then evaluated with the same raw microscopy images of these granules observed under polarized light, as has been previouly used for manual identification and counting. A dataset comprising 1308 2564 × 1924-pixel images is analysed. The images contain 17000 labelled granules of starch for two types of wheat: commercial wheat flour from 'Castilla' (type 0) and the Galician autochthonous flour 'Caaveiro' (type 1). The number of samples is approximately the same for each class. Instance segmentation with Mask R–CNN (Model II) achieved valid results for unseen images, with a categorical global accuracy of about 88.6% and with a discrepancy with respect to the areas of the granules as estimated by a human expert of less than 4%. The performance achieved by Mask R–CNN produces a strong correlation between the results of an expert and the results of the network, confirming the practical validity of our proposal.

1. Introduction

The European protected geographical indication (PGI) 'Galician Bread' (European Commission, 2019) has been recently obtained by a traditional bread from Galicia (NW Spain). This indication certifies that this traditional bread is made in part of commercial wheat flour (*Triticum aestivum* L.) but a minimum of 25% of the flour must come from the Galician autochthonous cultivar of wheat ('Caaveiro'). This wheat differs from other cultivars because it has a higher percentage of protein, but with a medium-low strength as well as a darker colour, resulting in a bread with more intense aromas and flavour.

Producing bread under the PGI 'Galician Bread' has several benefits for producers and consumers. On the one hand, it guarantees that the bread is made with quality ingredients and following strict rules of preparation and packaging. On the other hand, it allows consumers to differentiate the product from other types of bread and offers added

* Corresponding author.
E-mail address: xose.vidal@usc.es (X.R. Fdez-Vidal).

https://doi.org/10.1016/j.foodcont.2023.110198
Received 23 June 2023; Received in revised form 2 October 2023; Accepted 2 November 2023
Available online 8 November 2023

value to the consumer looking for a traditional, artisanal product with an authentic flavour. In order to guarantee its quality and authenticity, a regulatory board must verify that the rules established for the production and use of permitted flours with the stipulated percentages (a minimum of 25% of the flour must come from the Galician autochthonous cultivar) are followed. To avoid fraud, an automatic, quick, reliable and low cost method would be desirable to check the types of flour in a sample (common wheat flour and autochthonous cultivar), their percentages and the size and type of the starch using microscopy images.

Expert-supervised techniques can identify which cereal a flour comes from and its composition by using different techniques, such as chromatography, spectrometry, and spectroscopy –including Near-Infrared Spectroscopy (NIRS)– to identify its ingredients and additives and detect food fraud, either alone or in combination with other methods that allow the chemical separation of similar compounds in food (Knödler et al., 2010; Verdú et al., 2016; Ziegler et al., 2016). Another reliable method that is used to trace the geographical origin of wheat is isotope ratio analysis (Rashmi et al., 2017). Techniques based on molecular analysis, such as DNA-based methods, are also used to identify original wheat products (Morcia et al., 2020; Ramos-Cabrer et al., 2022).

These techniques involve complex, manual and time-consuming methods that must be carried out by specialised professionals to identify the varieties of flour that make up a mixture and to determine their percentages (Liu et al., 2023; Morcia et al., 2020). However, the method we propose to carry out a quick and reliable routine control of the flours is quite simple. It consists of taking photographs of the flours under polarized light in an optical microscope and then identifying the two types of starch by their morphology and counting them to ensure that the mixture has the minimum percentage of autochthonous flour required by the PGI. A previous publication (Fernández-Canto et al., 2023) described the manual method of starch identification and counting, but now we turn to artificial intelligence, using neural networks, to automate the process.

Our approach is supported by the recent work of Fernández-Canto et al. (2023), who developed a sample preparation method (minimising lumps) that takes advantage of the fact that when it is subjected to polarized light microscopy, the starch of 'Caaveiro' flour (Type 1) is morphologically different (Fig. 1b) from that of commercial wheat flour 'Castilla' (Type 0) (Fig. 1a). This can be used to determine the proportion of 'Caaveiro' in a flour blend by counting the proportion of granules of each type of starch. Due to this difference, a simple mathematical model has been developed that enables the quantification of the proportion of 'Caaveiro' in a mixture up to a detection level of 5.17% in a simple and very low-cost way. Through this work, the authors developed a microscopic technique to verify the presence of 'Caaveiro' flour in a mixture and to quantify its proportion, guaranteeing 25% of 'Caaveiro', the minimum quantity required in 'Galician Bread'. In order to expedite the process of discerning each type of starch (this counting was done manually), and to reduce the amount of tedious counting, we tried to automate the identification and counting of each type of starch through the use of computer vision techniques.

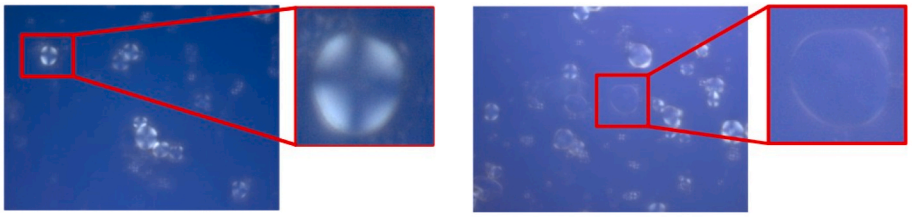
Intrinsically, once the microscopy digital photographs were taken, this process involves the detection, classification and segmentation of the granules of starch in the photographs of the flour samples. Our problem is similar to that encountered for granular materials composed of a large collection of tightly packed solid particles surrounded by a gas or liquid. Automatic processes must deal with different types of these materials, which require advanced techniques for measuring the parameters of their morphologies, including the number of particles, their sizes and shapes; additionally, an evaluation of the individual particle based on its texture, colour or other parameters must be performed to guarantee the quality and efficiency of the products and systems.

Techniques based on image processing for quality control are very popular in the agricultural and food industries. This type of approach has several advantages, including rapidity, low cost, and an adequate accuracy; it can be applied to grains/seeds (Aslan et al., 2017), rice (Mohanty et al., 2022), flours (Kurtulmuş et al., 2014) and other agricultural by-products. Although our problem is somewhat different, it is similar to the segmentation of granular materials and minerals (Karimpouli & Tahmasebi, 2019), nanoparticles (Okunev et al., 2020) and microscopic algae (Ruiz-Santaquiteria et al., 2020).

However, the traditional image processing methods cannot autonomously segment particles in images containing particles with a high level of contact and/or overlap. Typical methods for the detection of circular shapes under perspective transformations include filtering with radially symmetric templates and voting-based shape-matching methods. Nevertheless, this kind of blob detector has a low degree of specificity when it comes to detecting shapes, and the responses are not well localised.

Regarding the voting-based approach, it starts with contour extraction. Contour points vote on the possible locations of the centre of the shape, which are computed from a geometrical model of the curve sought. Strong local maxima of the response of these points are highly likely to correspond to the centre of the curve. Examples of this are the fast radial symmetry detector of Loy and Zelinsky (2003) and the rotation- and scaling-invariant Hough ellipse detector of Nixon and Aguado (2008). However, voting-based algorithms are not actually invariant to scale, given that the intensity of the response increases with the number of votes, i.e. the size of the contour. Moreover, each contour point, whether it is a target boundary or a noise artifact, votes for one or many possible uncertain locations of the centre, yielding a cluttered and imprecise response (see Fig. 2). Therefore, traditional standalone image processing methods are not able to obtain good results with images that have overlapping particles.

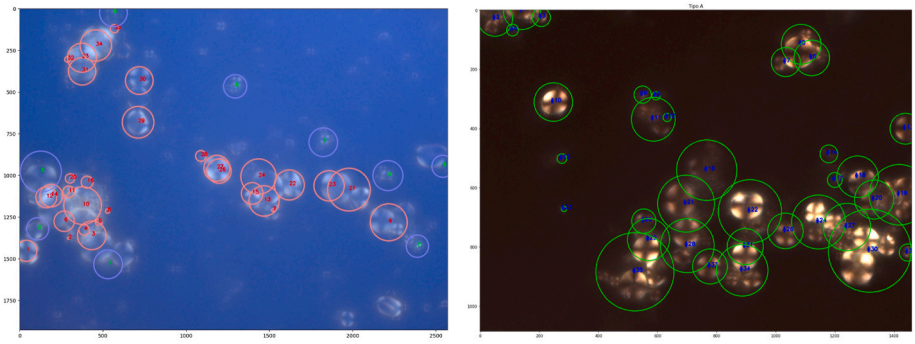
Currently, there are numerous variants of neural networks capable of performing instance segmentation (a combination of object detection and semantic segmentation methods), such as Mask R–CNN (He et al., 2017), Yolact (Bolya et al., 2019) and BlendMask (Chen et al., 2020), among others. These CNNs are making it possible to tackle very complex problems (occlusions, motion, etc.) and obtain performances far



(a) Commercial wheat flour 'Castilla' (Type 0)

(b) Autochthonous wheat flour 'Caaveiro' from Galicia (Type 1)

Fig. 1. Images under polarized light of the two types of flour we want to classify and count the granules of: (a) granules of starch of the commercial wheat flour 'Castilla' (type 0) whose main visual characteristic is the Maltese cross in a central position and (b) granules of starch of the 'Caaveiro' flour (type 1) which lacks the previous visual pattern and has a bright halo on its periphery.



(a) Case I (b) Case II

Fig. 2. Granule segmentation results obtained with classical computer vision techniques: (a) Hough's transform and (b) radial symmetry + watershed. As shown in the pictures, the traditional standalone image processing methods are not able to obtain good results with images that have overlapping particles, which is responsible for a cluttered and imprecise response.

superior to those of classical techniques.

The most significant advantage of CNNs that makes them superior to conventional methods is their ability to learn a large set of visual features, which allows the network to solve a problem without the need to preprocess the input image to detect the regions of interest (RoIs) and classify them (manual descriptor–classifier pair design). Deep CNNs have a large number of tunable parameters that allow them to learn hierarchically more and more complex features from the labelled data (ground truth) provided. Therefore, together with the possibility of knowledge transfer (learned in other tasks), they have gained great popularity in a wide variety of disciplines, such as civil engineering (Zhang, Yin, & Yan, 2022), face recognition (Li et al., 2020), action recognition (Hu et al., 2023) and medical image analysis (Soffer et al., 2019).

It is an efficient method for the detection and segmentation of objects in images that generates a high-quality mask for each instance of each type of flour. Therefore, it allows us not only to quantify the types of flour but also to determine the size and eccentricity of each granule.

In this paper, due to the good results it has achieved, outperforming the COCO 2015; COCO 2016 segmentation challenge winners, we propose to use a fine-tuned version of the Mask R–CNN method for our flour granule segmentation problem. Mask R–CNN is a deep learning model that combines object detection and instance segmentation within a single network. One of the advantages of Mask R–CNN is that it can handle particle overlap, i.e. when two or more objects partially or completely overlap. This is because Mask R–CNN uses a feature pyramid network (FPN) and a residual network (ResNet) to extract features at different levels of resolution and combine them into a single output. In contrast, classical methods rely on edge and geometric shape detection, which can fail when there is overlap or noise. Therefore, Mask R–CNN offers a higher accuracy and robustness for segmenting overlapping particles.

In a previous study, Fernández-Canto et al. (2023) demonstrated the validity of a new way of guaranteeing the identity of the type of wheat flour, using polarized light microscopy and validated by genetic analysis (Ramos-Cabrer et al., 2022). The main goal of this paper is to apply neural image processing techniques to automate the manual process of classification, segmentation and size determination to determine the proportion of 'Caaveiro' in a flour blend by counting the proportion of granules of each type of starch. The main contributions of this work are the following.

1. It demonstrates that the work of Fernández-Canto et al. (2023) is suitable for generating microscopy images that, under polarized light, allow the application of machine vision techniques for the automatic verification of compliance with the conditions of the PGI 'Galician Bread'.
2. A Mask R–CNN model is trained using fine-tuning to achieve an efficient method for the detection and identification of flour granules and the segmentation of these granules, which allows us to measure the granule morphology.
3. A new image dataset was built and labelled by an expert (one of the authors, N. F.-C.), with the following parameters: 1) the type of starch (type 0 – commercial wheat flour or type 1 – 'Caaveiro' flour) was labelled, and 2) a mask was assigned to each labelled granule to be segmented according to human specifications/criteria.

2. Materials and methods

2.1. Flour samples

The samples of flour were prepared as previously described in detail in Fernández-Canto et al. (2023). To recapitulate briefly, mixed flour samples were prepared from pure flours ('Caaveiro' and commercial 'Castilla') with decreasing percentages of 'Caaveiro' according to these ratios: 100:0, 90:10, 80:20, 70:30, 60:40, 50:50, 40:60, 30:70, 20:80, 10:90, 0:100. An additional mixture of 25:75 was included, as 25% of 'Caaveiro' is the minimum required to produce the PGI 'Galician Bread'. Subsequently, 10 mg of flour from each mixture were weighed, introduced into an eppendorf and 1 mL of gelatin solution was added and mixed homogeneously on a vortex mixer at 15 Hz for 5 min. Then, 200 μL of the suspension was distributed onto a microscope slide and allowed to dry at room temperature. Ten slides of each sample were prepared and the appearance, shape and size of the granules of starch were observed under light microscopy and under polarized light (as a result of adding polarising filters). From each slide, 15–20 digital photographs were taken with a Zeiss Axiophot microscope equipped with a digital camera (Zeiss AXIOCAM 208 color). Each photograph had a size of 2564 × 1924 pixels, in order to obtain a minimum of 400–450 granules per sample. Using the digital photographs and based on the birefringence observed in the granules of starch, a manual classification was performed in terms of two classes, type 0 and type 1, based on their different morphologies (see Fig. 1). Note that if the granules of starch of the different types of flour were to have the same morphology, this quantification method could not be carried out. All the photographs used for the manual quantification of each type of starch in the flour mixtures with different proportions were also used to train the neural network.

2.2. Mask R–CNN

The Mask R–CNN model (He et al., 2017) is an extension of the faster region-based convolutional neural network (Faster R–CNN) object detector (Ren et al., 2017). Mask R–CNN is based on two main

components: a feature pyramid network (FPN) and a residual network (ResNet). The FPN combines features at different scales to improve the detection and segmentation of objects at different scales. The ResNet provides a deep and robust architecture for extracting significant features from images. Mask R–CNN also uses a region proposal network (RPN) to generate candidate objects and classifier and regressor heads to predict their classes and locations (Fig. 3).

We used Mask R–CNN models with two different backbones: Model I has a ResNet-50-FPN backbone from He et al. (2017) and Model II has a ResNet-50-FPN backbone with visual transformers from Li et al. (2021). Both models are pre-trained on the COCO dataset (Lin et al., 2014), allowing them to learn generalisable features from a large dataset. Fine-tuning this pre-trained model to suit a specific task can significantly reduce the amount of time and resources required compared to training a new model from scratch. Specifically, there are three benefits of fine-tuning learning compared to a full ad hoc learning process: it is much faster (full learning can take weeks), it has a higher accuracy (once the model is tuned it can perform 20% better) and it needs less data to keep improving the model (pre-training allows it to detect specific features). In this case, we only need to modify the box predictor layer, classification layer and mask predictor layer to fit the custom dataset. For this, the Torchvision library (TorchVision maintainers and contributors, 2016) has been used to load the network architectures as well as the training reference scripts to perform the fine-tuning.

2.3. Data augmentation

That abundant data should be available for training a neural network is a crucial issue (Zhao, 2017). Thus, depending on the complexity of the task and the difficulty of obtaining and labelling the data, a large investment in terms of time and human resources may be necessary to obtain the amount of data required for successful training. Data augmentation is a technique that can be used to artificially create new training data from existing training data by applying domain-specific techniques to examples from the training data to create new and different training examples. This technique is often used in deep neural networks to improve their performance and invariance by increasing the amount of data available and its variability. The set of transformations was carried out with the Albumentations library (Buslaev et al., 2020). The selected transformations were brightness and contrast alteration, RGB band shifting, scaling, horizontal and vertical flipping, rotation and affine transformations with identical probability ($p = 0.4$).

2.4. Dataset and training

A key step in the training of a neural network is the construction and labelling of a good database. The properties we want are variability in the conditions under which the images were captured, balanced classes, and a sufficient quantity of images.

Our training and validation dataset consists of 1008 images (80% for training and 20% for validation) and the corresponding JSON file in COCO format (Lin et al., 2014) that contains the annotations of 14890 samples of flours labelled by the human expert. The samples of the flours were labelled as polygons using the MakeSense online Web tool (Fig. 4). The test dataset was prepared in the same way. It contains 300 images with 2111 labelled granules of starch and the corresponding COCO format files.

The training was carried out on a computer with a 16-core, 32-thread AMD® Ryzen threadripper 2950X processor and an NVIDIA® GeForce RTX 2080 Ti graphic accelerator. The networks were fine-tuned (classifier and FPN mask predictor heads) for 40 epochs (after 20 epochs, the network learning stabilises, as shown in Fig. 5), and a stochastic gradient descent (SGD) optimiser with an initial learning rate = 0.005, momentum = 0.9, weight decay = 0.0005 and batch size = 5 was used. We used a learning schedule to make the learning rate adapt to the gradient descent optimisation procedure (SGD), increase the performance, and reduce the training time. The learning rate of each parameter group decays by gamma = 0.1 every 3 epochs (step size).

2.5. Performance evaluation

To estimate the performance of our model, we adopt the following standard metrics.

1. **Performance metrics used to evaluate the classification performance**: To estimate the performance of our model, we adopt the

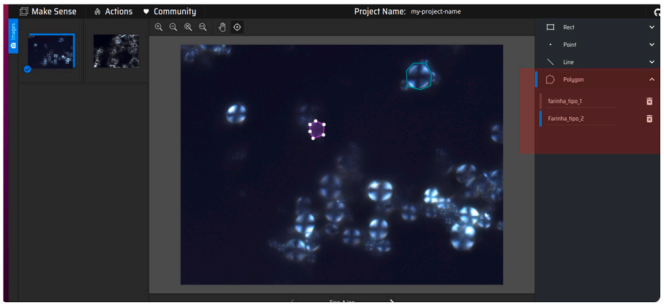


Fig. 4. Manual labelling by an expert of the granules of starch in each image of the dataset done using the MakeSense tool. These labelled images are called the *ground truth* and are used to estimate the performance of our networks.

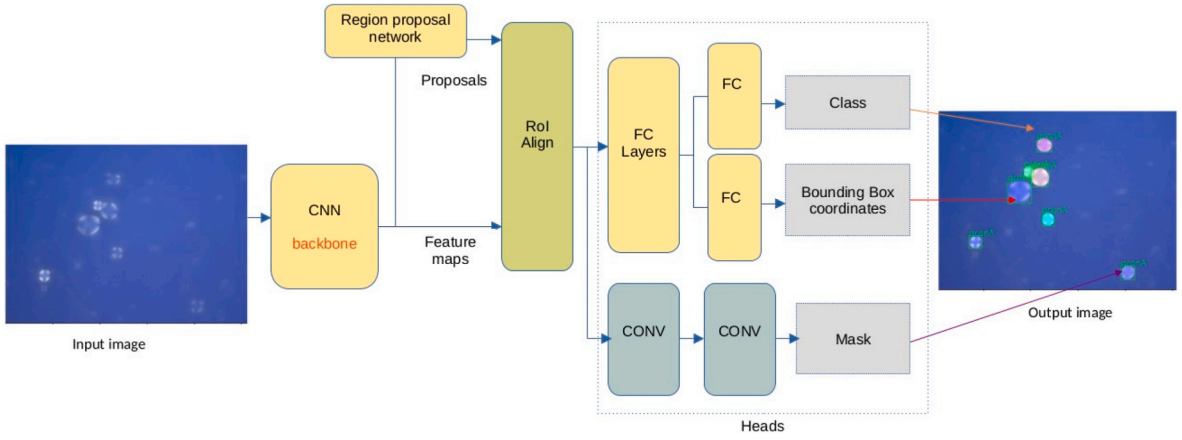


Fig. 3. Architecture of the Mask R–CNN model. The TorchVision library (TorchVision maintainers and contributors, 2016) provides the implementation of the architectures used in this work as well as the training reference scripts to perform fine-tuning.

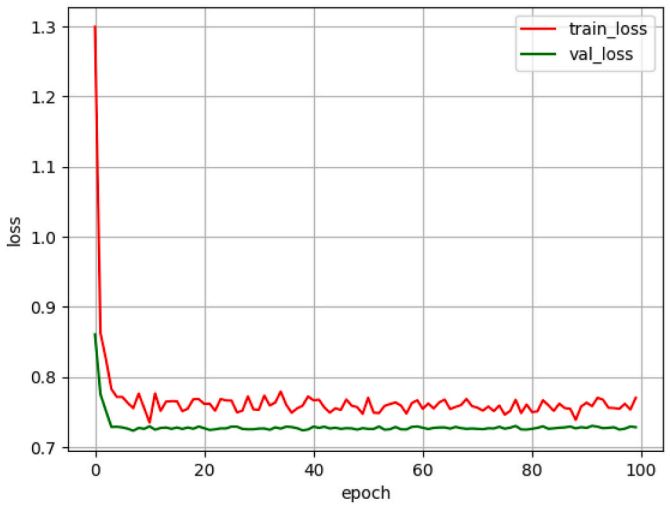


Fig. 5. Evolution of the loss measures for training and validation versus the number of epochs. The learning of the network is stable after 20 epochs, and does not improve significantly in the later iterations of training.

standard metrics: the accuracy, precision, recall and F_1 score. For more details, see Appendix A.

2. **Performance of the detection and instance segmentation**: Detection/instance segmentation models produce a collection of local bounding boxes/segmentation masks that describe each object that is detected/segmented in the image. As such, the methods of evaluating object detection and instance segmentation are quite similar, with the exception that we now calculate the metrics for masks instead of bounding boxes. To evaluate our collection of bounding boxes and predicted masks, we used the mean average precision (mAP) (Everingham et al., 2010) to measure the quality of the learned model in all categories. In our case, we used the mAP with a set of threshold values {0.55, 0.6, …0.95} and denote it by $\text{mAP}_{[0.5–0.95]}$. In object detection or instance segmentation, the objective is not only to correctly classify the flour in the image but also to find where in the image it is located; we cannot simply use the image classification metrics. In brief, the metrics estimate the extent of the intersection over union between the predictions and ground truth averaged over all classes, and the set of threshold values. Values of the mAP of 0% correspond to poor predictions, while an mAP close to 100% signals the pixel-to-pixel coincidence of the predictions and the ground truth. On the COCO database, the values achieved by the Cascade Mask R–CNN model are around 40% (Cai & Vasconcelos, 2018). The mAP metrics were evaluated using a COCO API tool (Dollar & Lin, 2020), along with annotations (annType) of the 'bbox' (detection) and 'segm' (mask segmentation) type and a maximum particle number (maxDets) of 100.

2.6. Postprocessing

While neural networks can produce predictions of the contours of the granules of the flour, it is important to have a well-defined algorithm to refine the contours of the mask predicted by the network. In this work, we implement the following postprocessing strategy, which consists of refining the contours of the masks predicted by the network by means of an elliptical fit (Halíř & Flusser, 1998). This approach has been proven to provide a good tradeoff between precision and computational cost compared to more sophisticated models. The effective size of a granule of flour was calculated as the area of the resulting elliptic fit, as shown in Fig. 7e.

3. Results and discussion

3.1. Samples of flour

We have developed a method for the identification and quantification of each type of wheat flour (commercial wheat 'Castilla' versus 'Caaveiro') based on a microscopic analysis of their granules of starch to estimate the percentage of each flour in a mixture. Previoulsy, this was done manually, and now we are trying to automate it by using neural networks.

Other methods used in the daily inspection of wheat flours include near-infrared spectroscopy (NIRS), which is a reliable tool that allows detecting any adulteration of common wheat flour (Zhang, Liu, et al., 2022): for safety –such as excessive talc powder illegally added into the flour, or even ashes (Deng et al., 2019; Dong & Sun, 2013)– or just for quality control when blended with other flours of lower quality.

NIRS combined with chemometric methods allows detecting adulteration of wheat flour, which includes adding flours made from different tubers (casava, potato …), grains of different species of cereals (sorghum, rice, corn …), or even different species of wheat, due to their different commercial value (e.g. *Triticum durum* –more expensive– versus *Triticum aestivum* (Unuvar et al., 2021)). However, to the best of the authors' knowledge, there is nothing in the literature about detecting percentages of different varieties or cultivars from the same species of wheat, as in our case, where both pure flour samples are *Triticum aestivum*.

3.2. Performance in classification, localisation and segmentation of the granules of starch

Tables 1 and 2 show the confusion matrices obtained by the Mask R–CNN models with different backbones (Models I and II) on the test database. To calculate the classification performance measures, human labelling was only performed for the most significant particles in the image, not for all the particles that the network is able to detect. To deal with this problem, the classification performance measures only take into account those particles detected by the system with a bijective relationship to those labelled by the human expert. This decision entails discarding the system's hits/misses for detected particles with no correspondence to the ground truth and vice versa. Due to this decision, the ground-truth support used for the calculation of the performance measures may vary slightly. The results achieved by Model II (Table 3), which has visual transformers, are slightly better than those of Model I (Table 3), but they are statistically equivalent, with an accuracy higher than 88%. Looking at the performance measures for each class separately, we see that the model is more accurate for the class type 0 than for the class type 1 (it minimises the false positives (FPs)). However, the recall values are higher for the class type 1 (it is able to recover 90.5% of the true positives (TPs) of this class while minimising the false negatives (FNs)). The overall F_1 score is 88.4%. The measures $\text{mAP}^{bbox}_{[0.5–0.95]} = 0.539$ and $\text{mAP}^{segm}_{[0.5–0.95]} = 0.547$, which are used to evaluate detection and segmentation, are significant and reveal a very good correspondence of the predicted boxes and masks (network) with their ground truth (human expert) counterparts.

In Fig. 6, we present a sample of the results obtained for a sample of images from the test database. Column a of Fig. 6 shows the original

Table 1
Confusion matrix for Model I.

		Hypothesis		
		Ω_0	Ω_1	$\sum_{rows}$
Reference	Ω_0	1063	91	1154
	Ω_1	152	792	944
	$\sum_{cols}$	1215	883	

Table 2
Confusion matrix for Model II.

	Hypothesis			
		Ω_0	Ω_1	$\sum_{rows}$
Reference	Ω_0	1071	82	1153
	Ω_1	156	785	941
	$\sum_{cols}$	1227	867	

image, column b shows the labels and the region of each granule (type 0 or type 1) entered manually by the expert (ground truth) and finally, column c shows the output of the neural network, which includes the predictions made by the network for each detected granule and the corresponding segmentation mask (displayed with random colours). This diagram permits a visual confirmation that the model's high quantitative evaluation values result in a strong agreement, in terms of classification, segmentation, and sizing, between the expert's labelling (column b) and the network's predictions (column c). We can see that the predicted segmentation fits the real contour of the different granules well. In Fig. 7b, an enlargement of a patch of an image from the database (Fig. 7a) is presented to appreciate the details in the results; in Fig. 7c, d and e, the segmentation mask produced manually by the human expert (ground truth), the network predictions, and the elliptical fitting over the points of each granule's contours segmented by the network output, respectively, are presented.

The overall statistical behaviour of the network is described by the performance measures shown above. However, it is difficult to determine from these values whether the network is a reliable replacement of a classification made by a human expert. To address this, 10 flour blends were created using pure flours with decreasing percentages of 'Caaveiro' according to the specified ratios: 75:25, 70:30, 65:35, 60:40, 55:45, 45:55, 40:60, 35:65, 30:70, 25:75. Ten images were taken for each mixture, and both the human expert and the neural network counted the granules. This gives us the average proportions of each type of granule for each mixture. Fig. 8 shows the real percentage of Type 0 flour on the x-axis for each experiment. On the y-axis, we plot the true percentage of Type 1 flour (blue line) and the predicted percentages for each experiment by both the human (yellow polygonal line) and Type 1 network (green polygonal line). The graph shows that there is a significant degree of agreement between the human and network predictions and very close to the true percentages. The average error for the human estimate compared to the true value is 1.6%, while for the network, it is 3.2%. The network's ability to learn and replicate human behaviour in the counting task, automating it with acceptable margins of error, can be confirmed as an initial proposal.

3.3. Influence of input image resolution on detection

The resolution of the input image has a significant influence on the detection process. To determine this influence, we took subsamples of the images of the test set with the factors {1, 0.5, 0.25, 0.125}, which reduces the size of the original image by up to one-eighth. For type 0 granules, the percentages of detected (ground truth) granules with respect to the rescaling factor are {94.6%, 93.8%, 93%, 91.2%}. From the morphology of the type 0 granule, we see that this transformation has hardly any influence on the detection. For type 1 granule, the percentages of detected ground-truth granules with respect to the re-scaling factor are {89.6%, 88.6%, 85.2%, 69.7%}. In this case, when the subsampling is high, the detection drops sharply because the removal of the high frequencies destroys the visual characteristics of the type 1 granules (see Fig. 1b).

3.4. Estimation of the elliptical area of granules

To quantify the results of the postprocessing method that determines the elliptical area of each granule using a contour fit extracted from the mask segmented by the network, we cannot use the polygonal approximations that the human expert marked to train the network. As can be seen in Fig. 7c, in most cases, the ground truth overestimates the real area. Therefore, we turned again to the human expert, instructing them to manually determine the ellipse that best fits each granule. To do this, we used the ImageJ (Schneider et al., 2012) tool, which allows us to obtain the value of the adjusted elliptical area (menu: Analyze → Measure). We randomly selected 10% of the test images and, in each one of them, we determined the ellipses that best fitted the granules of the two classes that the expert considered relevant; we tried to find a wide range of sizes (the number of samples of each type of flour was $N = 50$). To measure the discrepancy between the areas obtained by the model and those fitted by the human expert (there is no normality in the sample), we define the following measure of this discrepancy:

$$\mathscr{D} = \frac{1}{N}\sum_{i}^{N}\frac{|A_{i,Net} - A_{i,Hum}|}{A_{i,Hum}}, \tag{1}$$

where A_{Net} is the elliptical area estimate obtained from the postprocessing of each granule and A_{Hum} is the estimate made by the expert with the ImageJ tool. The final value of $\mathscr{D}$, for each type of flour, is normalised to the value of the area estimated by the human expert and averaged over the number of samples (N) of each flour. Table 4 shows the values of $\mathscr{D}$ and its standard deviation $\sigma(\mathscr{D}(\%))$ for the two types of flour.

The average discrepancy for type 0 granules is $\mathscr{D}_0 = 3.96\%$, and for type 1 granules, it is $\mathscr{D}_1 = 3.54\%$. The range of areas (pixels squared) sampled for type 0 granules is in the range [8763, 61557], and for type 1

Table 4
Mean particle sizes calculated from histograms of different types of contours.

Particle type	Discrepancies	
	$\mathscr{D}(\%)$	$\sigma(\mathscr{D}(\%))$
0	3.96	1.96
1	3.54	2.28

Table 3
Results for both models.

Model I (ResNet-50-FPN backbone)									
Classification								Detection and segmentation	
Class	For each class			Global				Global	
	Pr_κ	Re_κ	F_1	UAPr	UARe	F_1	ACC	$mAP^{bbox}_{[0.5--0.95]}$	$mAP^{segm}_{[0.5--0.95]}$
Ω_0	92.1%	87.5%	89.7%	**88%**	**88.6%**	**88.2%**	**88.4%**	**53.8%**	**53.5%**
Ω_1	83.9%	89.7%	86.7%						
Model II (ResNet-50-FPN backbone with visual transformers)									
Ω_0	92.9%	87.3%	90%	**88.2%**	**88.9%**	**88.4%**	**88.6%**	**53.9%**	**53.6%**
Ω_1	83.4%	90.5%	86.8%						

granules, it is in the range [8372, 10999]. The fits provided by the postprocessing method are valid, with average discrepancies less than 4%. By analysing the intra-class behaviour, we found that the system tends to underestimate the area of type 0 granules and, on the contrary, it tends to overestimate the area of type 1 granules.

Finally, Fig. 9 shows the histograms of the areas (squared pixels) obtained from the elliptical fit of each granule detected in the full test dataset. For the type 0 granules, the average area is $(2.1 \pm (1.0)) \times 10^4 pix^2$, and for the type 1 granules, it is $(2.2 \pm 1.1) \times 10^4 pix^2$. Using neural networks, our data are in line with those of Fernández-Canto et al. (2023) who stated that the granules of starch of commercial wheat 'Castilla' follow a bimodal distribution. According to Fernández-Canto et al. (2023) their diameter is small (<11.25 μm) or larger. In contrast, the diameter of granules of starch of 'Caaveiro' follows a trimodal distribution: less than 12.5 μm, medium size (12.5–22.5 μm) and larger size (>22.5 μm).

4. Conclusion

In this study, we have demonstrated that applying deep learning techniques to microscopy images of wheat flour samples, prepared using the method of Fernández-Canto et al. (2023), allowed us to automate the verification of the flours used to manufacture the PGI 'Galician Bread'. Instance segmentation with Mask R–CNN (Model II) achieved valid results for unseen images, with a categorical global accuracy of about 88.6%. The elliptic fits used to estimate the area are accurate, with a discrepancy with respect to the areas estimated by a human expert of less than 4%. However, there is still room for improvement in the performance of the network as the databases continue to grow and we embed the network in a continuous learning scheme. The performance measures mentioned above show a strong correlation between the results of an expert and the results of the network (see Fig. 6), confirming the practical validity of our proposal.

This method has some advantages, as it allows an automatic

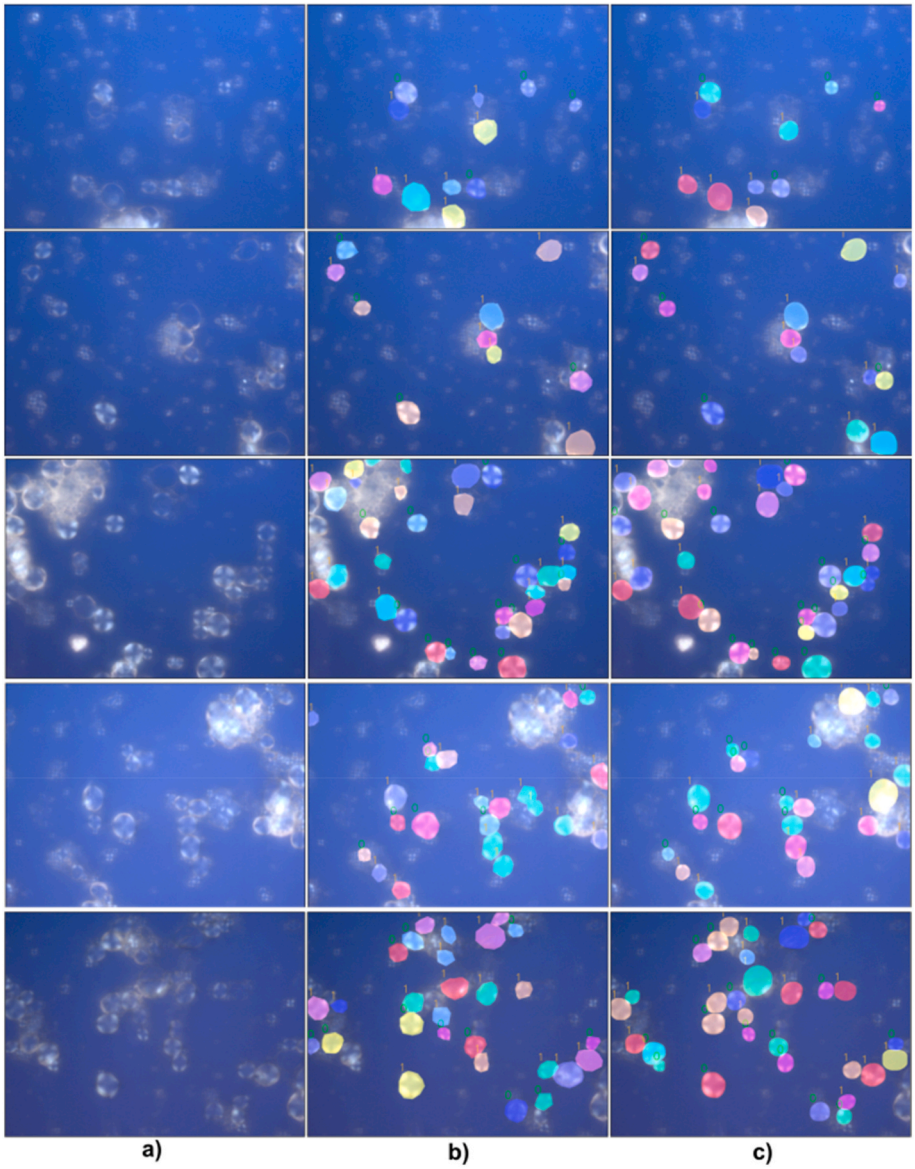


Fig. 6. A series of micrographs (test dataset) to show the qualitative results of the network predictions compared to the labelling performed by the human expert: column (a) input images, column (b) label and mask segmentation produced manually by human (ground truth), and column (c) label and mask segmentation predictions by the network. Each row corresponds to the same starch microscopy image. This diagram permits a visual confirmation that the model's high quantitative evaluation values result in a strong agreement, in terms of classification, segmentation, and sizing, between the expert's labelling (column b) and the network's output (column c).

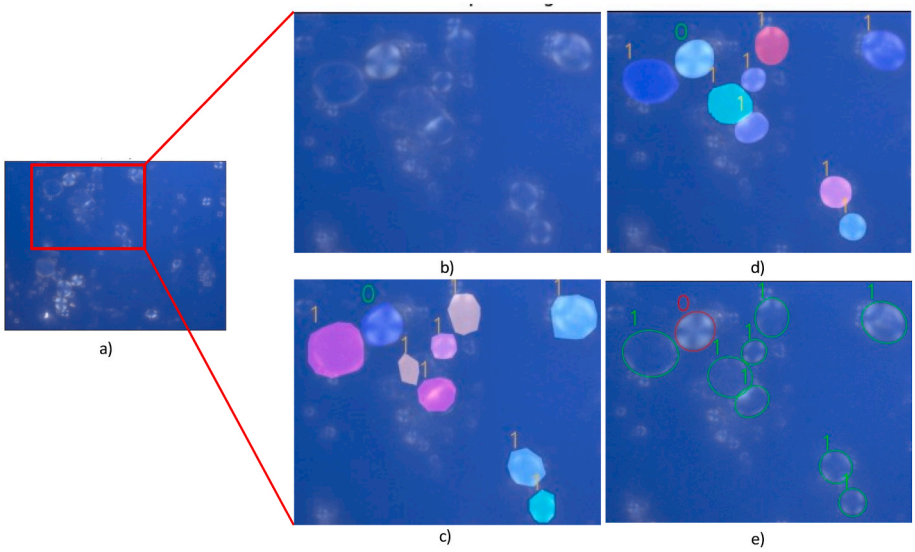


Fig. 7. Qualitative results of the segmentation quality and elliptic fitting: (a) input image, (b) original patch, (c) segmentation mask produced manually by human expert (ground truth), d) segmentation mask produced by the network and (e) ellipses obtained by an elliptical fit over the points of each granule's contours segmented by the network output (Fig. 7d).

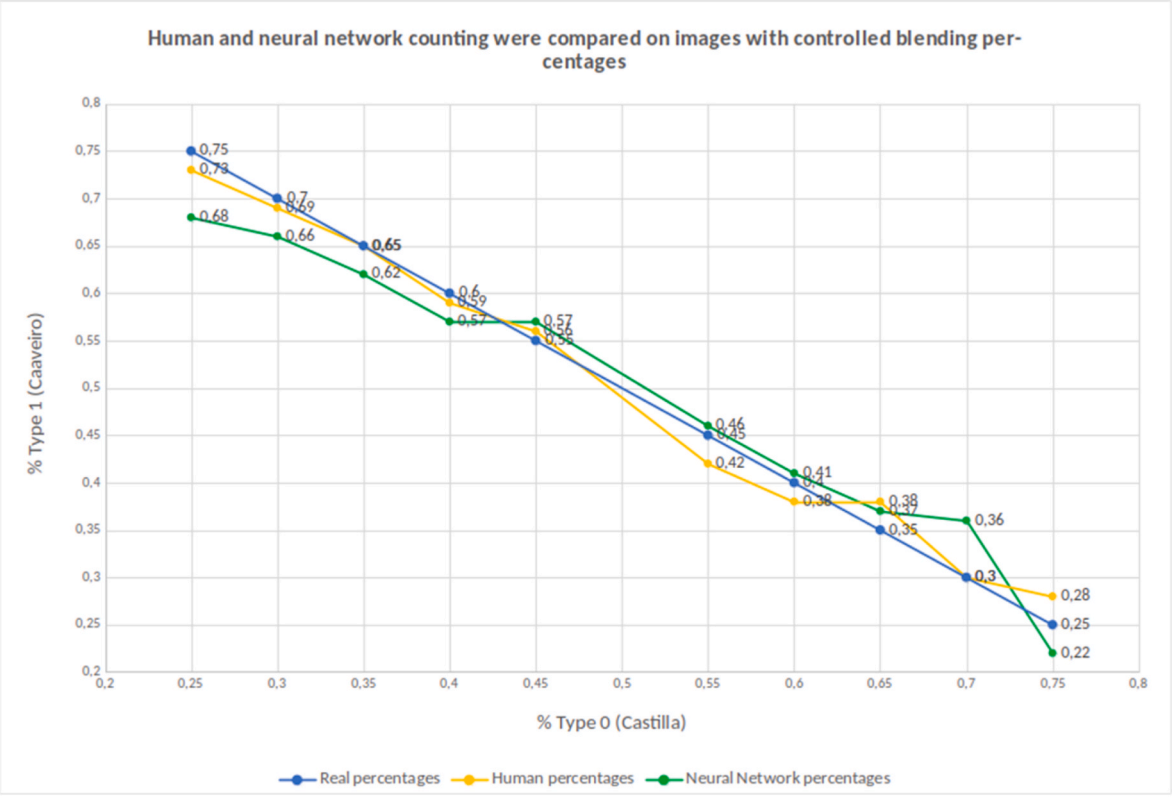


Fig. 8. Human and neural network counting compared on images with controlled blending percentages. Ten flour blends were created using pure flours ('Caaveiro' and 'Castilla') with decreasing percentages of 'Caaveiro' according to the specified ratios: 75:25, 70:30, 65:35, 60:40, 55:45, 45:55, 40:60, 35:65, 30:70, 25:75.

determination of the proportions of different kinds of flours within a blend. It could be applied as a quick and reliable method for quality control in the production of bread or in other flour mixtures (e.g. wheat starch versus tuber starch). However, it also has disadvantages, as this method can only be used for flour samples with starch granules morphologically different under polarised light, which is not the case for most of the wheat flour analysed.

For future work, the promising results of the Mask R–CNN model encourage us to continue working on two approaches: 1) controlling the perfectly balanced mixtures and increasing the number of granules of starch identified by the human expert to train the system; and 2) evaluating recent models that could allow us to achieve an instance segmentation of the granules of starch, such as YOLOv8 (Jocher et al., 2023) or a segment anything model (Kirillov et al., 2023) with YOLOv8 as its object detector.

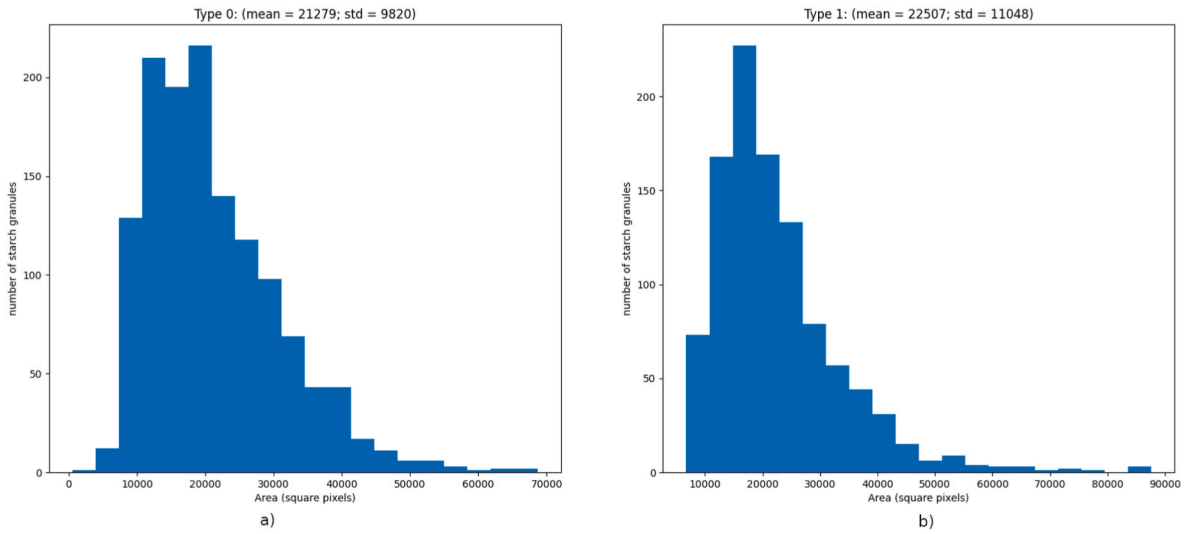


Fig. 9. Histograms of the areas estimated from the elliptical fit of the contours of the granules detected and segmented by the network (a) of type 0 and (b) type 1 for all test images.

Funding

This study was funded by the project 'Farm to Fork of Autochthonous Cereals in Ecological vs. Conventional Management' (CECOLECOPAN). Grants 2021 of "State Subprogram for Knowledge Generation" in the framework of State Program to Promote Scientific-Technical Research and its Transfer of the State Plan for Scientific and Technical Research and Innovation 2021–2023. PID2021-123905OB-I00.

Nerea Fernández-Canto is grateful to the Xunta de Galicia for her predoctoral research fellowship (ED481A-2019/263).

CRediT authorship contribution statement

Xosé R. Fdez-Vidal: Conceptualization, Formal analysis, Methodology, Visualization, Data curation, Writing – original draft, Writing – review & editing. **Nerea Fernández-Canto:** Methodology, Validation, Visualization. **María Ángeles Romero-Rodríguez:** Funding acquisition, Resources, Writing – review & editing. **Ana María Ramos-Cabrer:** Methodology, Validation. **Santiago Pereira-Lorenzo:** Funding acquisition, Formal analysis, Writing – review & editing. **Matilde Lombardero-Fernández:** Conceptualization, Project administration, Investigation, Methodology, Visualization, Data curation, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

The authors would like to thank the PhD student Mario López-Lombardero for his suggestion of using neural networks to automatically count and classify the granules of starch of our wheat flours. We also thank Fernando Almeida and Dacunha for providing the flour samples used in this study.

Appendix A. Performance metrics used to evaluate the classification performance

For a K-class problem with the confusion matrix given in Table A5, we will use the performance measures defined below.

Table A.5
Confusion matrix with absolute frequencies for a K-class problem.

	Hypothesis					
		Ω_1	Ω_2	…	Ω_K	$\sum$
Reference	Ω_1	n_{11}	n_{12}	…	n_{1K}	N_1
	Ω_2	n_{21}	n_{22}	…	n_{2K}	N_2
	⋮	⋮	⋮	⋱	⋮	⋮
	Ω_K	n_{K1}	n_{K2}	…	n_{KK}	N_K
	$\sum$					N

- *Accuracy*: The accuracy is the fraction of predictions that the model gets right. Formally, the accuracy has the following definition:

$$\mathrm{ACC} = \frac{1}{N}\sum_{\kappa=1}^{K} n_{\kappa\kappa}\cdot 100\,\%. \tag{A.1}$$

- *Precision*: The precision is calculated from the perspective of the prediction result for the class κ, and it is the proportion of true positive (TP) samples out of all the positive samples predicted by the model:

$$\mathrm{Pr}_{\kappa} = \frac{n_{\kappa\kappa}}{\sum_{i=1}^{K} n_{i\kappa}}. \tag{A.2}$$

- *(Unweighted) Average Precision*: The (unweighted) average precision is the average of Pr_{κ} for all K classes:

$$\mathrm{UAPr} = \frac{1}{K}\sum_{\kappa=1}^{K} \mathrm{Pr}_{\kappa}\cdot 100\,\%. \tag{A.3}$$

- *Recall*: The recall is calculated from the real sample set, and is the number of positive samples recovered for the class κ by the model divided by the total number of positive samples:

$$\mathrm{Re}_{\kappa} = \frac{n_{\kappa\kappa}}{\sum_{i=1}^{K} n_{\kappa i}}. \tag{A.4}$$

- *(Unweighted) Average Recall*: The (unweighted) average recall is the average of Re_{κ} for all K classes:

$$\mathrm{UARe} = \frac{1}{K}\sum_{\kappa=1}^{K} \mathrm{Re}_{\kappa}\cdot 100\,\%. \tag{A.5}$$

- F_1 *Score*: The F_1 score provides a single score that balances (harmonic mean) both the precision and recall in a single number. A good F_1 score means that the model has a low number of false positives (FPs) and a low number of false negatives (FNs). An F_1 score of 1 is considered perfect, while the model is a total failure if the F_1 score is 0:

$$F_1 = 2\cdot\frac{\mathrm{Re}\cdot\mathrm{Pr}}{\mathrm{Re}+\mathrm{Pr}}. \tag{A.6}$$

References

Aslan, M. F., Sabanci, K., Yiğit, E., Kayabasi, A., Toktas, A., & Duysak, H. (2017). A comparative classification of wheat grains for artificial neural network and extreme learning machine. *International Journal of Engineering Technologies IJET*, 231–236.

Bolya, D., Zhou, C., Xiao, F., & Lee, Y. J. (2019). Yolact: Real-time instance segmentation. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 9157–9166). https://doi.org/10.48550/arXiv.1904.02689

Buslaev, A., Iglovikov, V. I., Khvedchenya, E., Parinov, A., Druzhinin, M., & Kalinin, A. A. (2020). Albumentations: Fast and flexible image augmentations. *Information, 11*. https://doi.org/10.3390/info11020125

Cai, Z., & Vasconcelos, N. (2018). Cascade r-cnn: Delving into high quality object detection. In *2018 IEEE/CVF conference on computer vision and pattern recognition* (pp. 6154–6162). https://doi.org/10.1109/CVPR.2018.00644

Chen, H., Sun, K., Tian, Z., Shen, C., Huang, Y., & Yan, Y. (2020). Blendmask: Top-down meets bottom-up for instance segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 8573–8581). https://doi.org/10.48550/arXiv.2001.00309

Deng, J., Wang, W., Zhao, X., & Lu, Y. (2019). Near infrared multispectral detection of talc content in flour. *Modern Food Science and Technology, 11*, 270–276. https://doi.org/10.13982/j.mfst.1673-9078.2019.11.037

Dollar, P., & Lin, T. (2020). Coco api-dataset. https://github.com/cocodataset/cocoapi. (Accessed 1 June 2020).

Dong, X., & Sun, X. (2013). A case study of characteristic bands selection in near-infrared spectroscopy: Nondestructive detection of ash and moisture in wheat flour. *Journal of Food Measurement and Characterization, 7*, 141–148. https://doi.org/10.1007/s11694-013-9149-0

European Commission. (2019). Regulation 2019/2182 of 16 december 2019 entering a name in the register of protected designations of origin and protected geographical indications [pan galego (PGI)]. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019R2182&from=ES.

Everingham, M., Van Gool, L., Williams, C., Winn, J., & Zisserman, A. (2010). The pascal visual object classes (voc) challenge. *International Journal of Computer Vision, 88*, 303–338. https://doi.org/10.1007/s11263-009-0275-4

Fernández-Canto, N., Ángeles Romero-Rodríguez, M., Ramos-Cabrer, A. M., Pereira-Lorenzo, S., & Lombardero-Fernández, M. (2023). Polarized light microscopy guarantees the use of autochthonous wheat in the production of flour for the Protected Geographical Indication 'Galician Bread'. *Food Control, 147*, Article 109597. https://doi.org/10.1016/j.foodcont.2022.109597

Halíř, R., & Flusser, J. (1998). Numerically stable direct least squares fitting of ellipses. In *Proc. 6th international conference in central europe on computer graphics and visualization* (pp. 125–132). Citeseer: WSCG.

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. *Proceedings of the IEEE international conference on computer vision*, 2961–2969. https://doi.org/10.48550/arXiv.1703.06870

Hu, K., Jin, J., Zheng, F., Weng, L., & Ding, Y. (2023). Overview of behavior recognition based on deep learning. *Artificial Intelligence Review*, 1833–1865. https://doi.org/10.1007/s10462-022-10210-8

Jocher, G., Chaurasia, A., & Qiu, J. (2023). YOLO by ultralytics. URL: https://github.com/ultralytics/ultralytics.

Karimpouli, S., & Tahmasebi, P. (2019). Segmentation of digital rock images using deep convolutional autoencoder networks. *Computers & Geosciences, 126*, 142–150. https://doi.org/10.1016/j.cageo.2019.02.003

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W. Y., Dollár, P., & Girshick, R. (2023). *Segment anything*. arXiv:2304.02643.

Knödler, M., Most, M., Schieber, A., & Carle, R. (2010). A novel approach to authenticity control of whole grain durum wheat (triticum durum desf.) flour and pasta, based on analysis of alkylresorcinol composition. *Food Chemistry, 118*, 177–181. https://doi.org/10.1016/j.foodchem.2009.04.080

Kurtulmuş, F., Gürbüz, O., & Değirmencioğlu, N. (2014). Discriminating drying method of tarhana using computer vision. *Journal of Food Process Engineering, 37*, 263–271. https://doi.org/10.1111/jfpe.12092

Li, L., Mu, X., Li, S., & Peng, H. (2020). A review of face recognition technology. *IEEE Access, 8*, 139110–139120. https://doi.org/10.1109/ACCESS.2020.3011028

Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft COCO: Common objects in context. In D. Fleet, T. Pajdla, B. Schiele, & T. Tuytelaars (Eds.), *Computer vision – ECCV 2014* (pp.

740–755). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-10602-1_48.

Liu, H. Y., Wadood, S. A., Xia, Y., Liu, Y., Guo, H., Guo, B. L., & Gan, R. Y. (2023). Wheat authentication: An overview on different techniques and chemometric methods. *Critical Reviews in Food Science and Nutrition, 63*, 33–56. https://doi.org/10.1080/10408398.2021.1942783

Li, Y., Xie, S., Chen, X., Dollar, P., He, K., & Girshick, R. (2021). *Benchmarking detection transfer learning with vision transformers*. https://doi.org/10.48550/arXiv.2111.11429

Loy, G., & Zelinsky, A. (2003). Fast radial symmetry for detecting points of interest. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 25*, 959–973. https://doi.org/10.1109/TPAMI.2003.1217601

Mohanty, S. P., Kumar, A., Singh, R. K., Singh, A. K., & Singh, O. N. (2022). Deep learning based high-throughput phenotyping of chalkiness in rice exposed to high night temperature. *Plant Methods, 18*, 1–14. https://doi.org/10.1186/s13007-022-00839-5

Morcia, C., Bergami, R., Scaramagli, S., Ghizzoni, R., Carnevali, P., & Terzi, V. (2020). A chip digital pcr assay for quantification of common wheat contamination in pasta production chain. *Foods, 9*, 911. https://doi.org/10.3390/foods9070911

Nixon, M. S., & Aguado, A. S. (2008). *Feature extraction and image processing*. Academic Press.

Okunev, A., Mashukov, M., Nartova, A., & Matveev, A. (2020). Nanoparticle recognition on scanning probe microscopy images using computer vision and deep learning. *Nanomaterials, 10*, 1285. https://doi.org/10.3390/nano10071285

Ramos-Cabrer, A. M., Fernández-Canto, N., Almeida-García, F., Gorostidi, A., Lombardero-Fernández, M., Romero-Rodríguez, M. A., & Pereira-Lorenzo, S. (2022). Traceability of the local cultivar 'Caaveiro' in flour mixtures used to produce Galician bread by simple sequence repeats and droplet digital polymerase chain reaction technology. *International Journal of Food Science and Technology, 57*, 7085–7098. https://doi.org/10.1111/ijfs.16048

Rashmi, D., Shree, P., & Singh, D. K. (2017). Stable isotope ratio analysis in determining the geographical traceability of indian wheat. *Food Control, 79*, 169–176. https://doi.org/10.1016/j.foodcont.2017.03.025

Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 39*, 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031

Ruiz-Santaquiteria, J., Bueno, G., Deniz, O., Vallez, N., & Cristobal, G. (2020). Semantic versus instance segmentation in microscopic algae detection. *Engineering Applications of Artificial Intelligence, 87*, Article 103271. https://doi.org/10.1016/j.engappai.2019.103271

Schneider, C., Rasband, W., & Eliceiri, K. (2012). Nih image to imagej: 25 years of image analysis. *Nature Methods, 9*. https://doi.org/10.1038/nmeth.2089

Soffer, S., Ben-Cohen, A., Shimon, O., Amitai, M. M., Greenspan, H., & Klang, E. (2019). Convolutional neural networks for radiologic images: A radiologist's guide. *Radiology, 290*, 590–606. https://doi.org/10.1148/radiol.2018180547

TorchVision maintainers and contributors. *TorchVision: PyTorch's Computer Vision library*, (2016). URL: https://github.com/pytorch/vision.

Unuvar, A., Boyaci, I., & Koksel, H. (2021). A novel approach for rapid discrimination of common and durum wheat flours using spectroscopic analyses combined with chemometrics. *Journal of Cereal Science, 100*, Article 103269. https://doi.org/10.1016/j.jcs.2021.103269

Verdú, S., Vásquez, F., Grau, R., Ivorra, E., Sánchez, A. J., & Barat, J. M. (2016). Detection of adulterations with different grains in wheat products based on the hyperspectral image technique: The specific cases of flour and bread. *Food Control, 62*, 373–380. https://doi.org/10.1016/j.foodcont.2015.11.002

Zhang, S., Liu, S., Shen, L., Chen, S., He, L., & Liu, A. (2022a). Application of near-infrared spectroscopy for the nondestructive analysis of wheat flour: A review. *Current Research in Food Science, 5*, 1305–1312. https://doi.org/10.1016/j.crfs.2022.08.006. URL: https://www.sciencedirect.com/science/article/pii/S266592712200123X.

Zhang, Z., Yin, X., & Yan, Z. (2022b). Rapid data annotation for sand-like granular instance segmentation using mask-rcnn. *Automation in Construction, 133*, Article 103994. https://doi.org/10.1016/j.autcon.2021.103994

Zhao, W. (2017). Research on the deep learning of the small sample data based on transfer learning. *AIP Conference Proceedings, 1864*. https://doi.org/10.1063/1.4992835

Ziegler, J. U., Leitenberger, M., Longin, C. F. H., Würschum, T., Carle, R., & Schweiggert, R. M. (2016). Near-infrared reflectance spectroscopy for the rapid discrimination of kernels and flours of different wheat species. *Journal of Food Composition and Analysis, 51*, 30–36. https://doi.org/10.1016/j.jfca.2016.06.005