Corn kernel classification from few training samples

Patricia L. Suárez [a,*], Henry O. Velesaca [a], Dario Carpio [a], Angel D. Sappa [a,b]

[a] *ESPOL Polytechnic University, CIDIS, Campus Gustavo Galindo, Guayaquil 09-01-5863, Ecuador*
[b] *Computer Vision Center, Universitat Autònoma de Barcelona, 08193-Bellaterra, Barcelona, Spain*

ABSTRACT

This article presents an efficient approach to classify a set of corn kernels in contact, which may contain good, or defective kernels along with impurities. The proposed approach consists of two stages, the first one is a next-generation segmentation network, trained by using a set of synthesized images that is applied to divide the given image into a set of individual instances. An ad-hoc lightweight CNN architecture is then proposed to classify each instance into one of three categories (ie good, defective, and impurities). The segmentation network is trained using a strategy that avoids the time-consuming and human-error-prone task of manual data annotation. Regarding the classification stage, the proposed ad-hoc network is designed with only a few sets of layers to result in a lightweight architecture capable of being used in integrated solutions. Experimental results and comparisons with previous approaches showing both the improvement in accuracy and the reduction in time are provided. Finally, the segmentation and classification approach proposed can be easily adapted for use with other cereal types.

© 2023 The Authors. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co., Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

According to the FAO, world grain trade is forecast to expand to a new record in 2021/2022. Corn continues to predominate in the world cereal market.[1] Based on the shown statistics and the importance of this cereal in the world economy, it is imperative to define standards in its commercialization processes and quality analysis is one of them. At present, the quality analysis of corn kernels is carried out manually by experts (e.g., Paulsen et al. (2019)), however, this task is tedious, complex, and prone to errors, since there is a visual and subjective inspection process.

As mentioned above, it is necessary to define policies and standards that allow fair trade (e.g., Bacon (2005), Naylor (2017), Wilson (2010)); subjectivity should be eliminated from the quality analysis process. In recent years some approaches have been proposed for the analysis of the quality of different types of kernels. In general, these new approaches are based on the usage of deep learning techniques (e.g., Velesaca et al. (2020), Shamim et al. (2020), Singh and

Chaudhury (2020), Li et al. (2019)) on visible spectra images (i.e., RGB images). In addition to these approaches, a few contributions working on other spectral bands (e.g., ultraviolet near-infrared) have been also proposed by Cheng et al. (2019). Although interesting results have been obtained, as will be presented in the next section, most of them impose constraints in the given sample set (e.g., non-touching kernels) making difficult their use in real scenarios. The main drawback of deep learning-based techniques is related to the requirement of a large number of labeled images to obtain a reliable and generalizable solution. In general, image annotation is manually performed becoming a time-consuming task and sometimes a bottleneck in the whole process. To overcome this problem some authors have proposed to use synthesized datasets (e.g., Toda et al. (2020), Kar et al. (2019)), which allow the generation of large annotated datasets, as large as required by the training process.

This paper presents a technique for classifying the content of a given sample of corn grains into three categories: good grains, defective grains, and impurities. Good grains are those that meet the desired quality standards and are suitable for consumption or processing; defective corn grains include broken samples or samples infected by fungi or bacteria, which makes them undesirable for consumption or processing; finally, impurities are foreign materials that are found in the given sample, they can be organic (e.g., pieces of grass or other cereals) or inorganic (e.g., pebbles). The classification problem is tackled without imposing constraints on the distribution of the grains or the presence of impurities. The proposed approach consists of two stages,

---

instance segmentation, and classification, as normally proposed in these types of solutions Velesaca et al. (2021b). In the first stage, a novel strategy is proposed to avoid the time-consuming data annotation process. In the second stage, an efficient and lightweight architecture is proposed, which overcomes state-of-the-art corn kernel classification approaches The core idea is to implement an independent lightweight classification module that may be used independently in further applications. The manuscript is organized as follows. Section 2 presents works related to the segmentation and classification problems. Section 3 presents the approach proposed for classifying elements from the given sample set, together with a summary of the dataset generated for the current work. Experimental results and comparisons with different approaches are given in Section 4. Finally, conclusions are presented in Section 5.

2. Related works

As described above, this paper presents an approach to perform corn kernel segmentation and classification. This technique addresses the inspection of the quality of corn samples, which could contain good kernels, defective kernels, and impurities. In this section, state-of-the-art algorithms in these topics are reviewed. Firstly, the most relevant segmentation approaches generally used in the detection step, are summarized; secondly, classification techniques related to classical and deep learning-based approaches, are presented. Table 1 presents a summary of all literature reviewed in this section.

2.1. Segmentation techniques

Instance segmentation has been largely studied in the computer vision community for different topics. Regarding its usage in the segmentation of sets of cereal kernels, there are studies based on the usage of threshold to split up the given image into background and foreground; these techniques work with grayscale images and perform

binarization using a specific threshold value, which depends on the type of grain analyzed and the color of the used background (e.g., Paliwal et al. (2003), Kaur and Singh (2015), Arboleda et al. (2018)). More elaborated approaches have been also considered, for instance, the watershed technique has been used to segment grains from the given image; this technique uses grayscale images where the variations in tonality represent a topographic surface and the highest intensity values are the peaks while the lowest values are the valleys. The process begins, filling each valley with a different color, then continues to fill until adjacent regions begin to touch and the boundaries between each region are well defined. The result of this process is the image segmented in each region with different colors. In Huang et al. (2019) and Shrestha et al. (2016) the authors propose to use the watershed method to obtain the segmented instance of corn and wheat kernels respectively.

Like in most computer vision tasks, recently convolutional neural networks (CNN) are used for the instance segmentation task overcoming state-of-the-art approaches (e.g., Aukkapinyo et al. (2019), Ganesh et al. (2019), Toda et al. (2020), Velesaca et al. (2020)). Among the different proposals, the Mask R-CNN architecture He et al. (2017) has become a benchmark in the area of object detection and instance segmentation; this architecture extends the Faster R-CNN object detection framework Ren et al. (2016) by adding a branch for the generation of the masks at the end of the model, thus achieving the segmentation of instances for each output bounding box. Regarding the usage of the aforementioned approaches in the cereal kernel segmentation, in Toda et al. (2020) the authors propose to use the Mask R-CNN network to determine the morphological phenotype of barley grains. For the training phase, the authors use a dataset where each image is generated by the collage of real corn kernels. In general, the model trained with these synthesized images has given better results compared to the training with manually annotated real-world images. The strategy is validated using other types of grains such as oats, rice, lettuce, and wheat. Another work to perform the instance segmentation of corn kernels is presented

Table 1
Summary of literature reviewed in Section 2 grouped by type of grain—classification accuracy corresponds to the values given by the authors in their experimental result sections.

Kernel	Author(s)	Segmentation	Classification	Classif. % Acc.
	Javanmardi et al. (2021)	Multi-Threshold	CNN	98.10
	Velesaca et al. (2021a)	Mask R-CNN	–	–
	Zhang et al. (2021a)	Otsu	CNN	93.30
	Zhang et al. (2021b)	Otsu	CNN	97.50
	Adke et al. (2020)	Mask R-CNN	Mask R-CNN	97.50
	Ali et al. (2020)	Thresholding	Multilayer Perceptron	98.93
	Velesaca et al. (2020)	Mask R-CNN	CNN (CK-CNN)	95.60
	Cheng et al. (2019)	–	Random Forest	95.00
	Effendi et al. (2019)	Thresholding	K-NN	90.00
	Huang et al. (2019)	Watershed	CNN (GoogLeNet)	95.00
	Li et al. (2019)	Line Profile Segmentation Method	Maximum Likelihood Estimator	85.72
	Ni et al. (2019)	Background removal	ResNet	98.20
Corn	Qiu et al. (2019)	–	PLS-DA	99.19
	Ribeiro and Yao (2019)	–	LBP + PCA	99.00
	Daskalov et al. (2018)	–	SVM	93.05
	Ni et al. (2018)	Thresholding + Morphological Operations	CNN	97.00
	Qiu et al. (2018)	–	Fourier Transform Near-Infrared Spectroscopy	98.70
	Wu et al. (2018)	Thresholding	Support Vector Machine-Grid Search	97.44
	Zhu et al. (2018)	Thresholding	K-SVD method	94.00
	Ambrose et al. (2016b)	–	PLS-DA	95.00
	Ambrose et al. (2016a)	Thresholding	PLS-DA	95.60
	Williams and Kucheryavskiy (2016)	Thresholding	Object-Wise Method	93.00
	Yuan et al. (2016)	–	SVM	91.11
	Yang et al. (2015)	Background segmentation	SVM	97.25
	Zhu et al. (2015)	–	K-NN	92.67
Barley	Toda et al. (2020)	Mask R-CNN	Mask R-CNN	95.00
	Paliwal et al. (2003)	Thresholding	ANN	96.00
	Shao et al. (2021)	Watershed	Localization-based Counting Fully CNN	89.88
Rice	Aukkapinyo et al. (2019)	Marker-based watershed	Mask R-CNN	81.00
	Kaur and Singh (2015)	Grayscale thresholding	Geometric features	97.21
Wheat	Shrestha et al. (2016)	Watershed	ANN	72.80
Coffee	Arboleda et al. (2018)	–	Geometric features	100

in our previous work Velesaca et al. (2020). In this work, the segmentation is carried out through the Mask R-CNN network that uses the default parameters of the framework. Also, ResNet 101 is used as a backbone, and the pre-trained weights from the COCO dataset are considered. The segmentation approach has been fed with a set of real-world images manually annotated with the open collaboration tool Labelbox[2]; this tool allows drawing off the contour of each element (i.e., corn kernels and impurities) present in the given image. One of the main problems of this pipeline is to label a large dataset; it is a time-consuming task that can be affected by wrong annotations. More recently, trying to avoid this time-consuming task, we propose a novel strategy that generates synthesized images of clusters of corn kernels Velesaca et al. (2021a). A dataset, as large and diverse as required is generated, thus reducing the use of resources.

Contrary to previous works where clusters of corn kernels are considered, in Adke et al. (2020), the instances segmentation of corn ears are tackled to estimate the total consumption of the grains by animals in the wild. The authors use the Mask R-CNN network to address two approaches to instance segmentation: identifying whole corn ears and bare cob parts with and without corn kernels. It can be highlighted that the results obtained by the Mask R-CNN network when evaluated in images with different lighting conditions and backgrounds with which it was not trained, the network was able to identify certain instances of segmentation with acceptable precision. The training set contains 70 manually annotated images, although this is not a large number it should be highlighted that the contour of every single kernel in the ear should be drawn, hence it is a tedious and time-consuming task.

In Shao et al. (2021), similar to the previous work, the author performs the segmentation of grains on the ears, but with rice kernels, it uses the LC-FCN semantic segmentation network to perform the segmentation of rice ears, then applies the Watershed algorithm Roerdink and Meijster (2000) to obtain all individual instances present in the image. The experimental results obtained by the authors show better results than traditional machine learning methods and SSD object detection algorithms. The model allows the calculation of the number of ears of rice in the field and provides basic data to estimate the yields of rice production. Like the previous work, it has the problem that the images were manually labeled, which represents a time-consuming task. The Watershed segmentation technique used by the authors is very sensitive to lighting conditions, resulting in an over-segmentation and generation of a greater number of instances than there are.

2.2. Classification techniques

Once every single instance from the given image is segmented, they need to be classified according to the categories defined in the final application. Recently, some approaches for the corn kernel classification problem have been proposed, especially focused on determining the quality of the grain to minimize time and costs in the verification processes. For instance, in Cheng et al. (2019) the authors propose a pattern recognition-based technique to detect contaminated corn kernels using a back-propagation neural network; the main limitation of this approach lies in the fact that kernels should not touch each other. Also, there is another technique based on the Fourier transform using near-infrared images, with which the most representative characteristics were extracted to evaluate their quality Qiu et al. (2018). Another work focused on the corn kernel classification problem has been presented in Qiu et al. (2019), where the authors introduce a corn classification algorithm that extracts the most representative features to differentiate between good and defective corn kernels. The success of this approach depends on a pre-processing and normalization process. Another approach that performs corn classification of three types of

grades (A, B, C) is presented by Wu et al. (2018). This technique is based on a support vector machine to perform the classification. In Ali et al. (2020) the authors also present a corn seed classification technique using hybrid features to determine the corn seed quality. This approach has been developed using Random Forest Breiman (2001), BayesNet Pearl (1985), LogitBoost Friedman et al. (2000) an MLP model, which reaches the best results.

In Zhu et al. (2018), an approach to classify 4 types of corn kernels has been proposed using a learning model based on a multi-kernel dictionary. Similarly, in Effendi et al. (2019), a traditional classification method based on the KNN algorithm that extracts color and texture has been proposed to determine the quality of the corn grains. In Ambrose et al. (2016b) the authors propose an approach to determine the viability of seed germination, for which, they have implemented the use of techniques based on Raman spectroscopy and Fourier transform using images of the near-infrared spectrum. Similarly, in Williams and Kucheryavskiy (2016) an approach is proposed for the classification of corn grains according to the hardness of the grain. For the experiments, images of the hyperspectral near-infrared spectrum have been used based on the object-wise algorithm. Another technique to determine the purity of the corn seed is proposed in Yang et al. (2015), where the corn seed classifier is based on texture, spectral, and morphological characteristics using SVM. The authors do not indicate whether the samples used in their experiments include seeds with occlusion or that are touching each other, which would be important to consider to see if it works with partial parts of the images of grains or with undefined edges that make look like a grain has two germ core.

In Zhu et al. (2015) a method is proposed to classify maize kernels contaminated with aflatoxins using the KNN algorithm. For this work, hyperspectral images of the near-infrared spectrum have been used. In Yuan et al. (2016), the authors propose a technique to detect moldy corn kernels with different levels of mildew using a support vector machine and images of the near-infrared spectrum. In Daskalov et al. (2018), the authors also present a corn kernel classification algorithm to detect or rule out the presence of kernels affected with Fusarium moniliforme disease by implementing a support vector machine. Also using images beyond the visible spectrum, in Ambrose et al. (2016a) the authors propose a hyperspectral imaging technique to classify corn seeds; the proposed approach receives the characteristics extracted from partial least squares discriminant analysis (PLS-DA) Wold et al. (2001) for the grading of aged (heat-treated) and normal (untreated) corn seeds. In this work, the authors propose a technique that is not invariant to the multiple positions of the grains, in the experiments they are only based on the diagonal and intercalated position of the seeds. Also, in Ribeiro and Yao (2019), the authors propose a combined HSV color space image processing technique to classify corn kernels between good and rotten. The authors describe that their data set has been based on the same set of images at three different resolutions.

Contrary to previous approaches, which are based on traditional machine learning techniques, in Zhang et al. (2021a) a CNN-based approach has been proposed to perform four varieties of corn seed classification, to determine the quality of the corn kernels. The model works with hyperspectral images of certain wavelength ranges, corresponding to the seed zone. The results obtained by the convolutional network were better when compared with techniques such as K-nearest neighbors Cover and Hart (1967) and support vector machine Cortes and Vapnik (1995). On the other hand, in Li et al. (2019), the authors propose a deep learning method to differentiate defective classes of corns. Although interesting results are presented, it is not clear what is the behavior of the model with touching or occluding grains. In the same way, Ni et al. (2018) proposes deep convolutional networks, such as AlexNet Krizhevsky et al. (2012) and VGG Simonyan and Zisserman (2015), to classify dual touching kernels. Another proposal for the classification of grains is presented in Zhang et al. (2021b), where the authors propose the classification of corn grains by their

---

[2] labelbox.com.
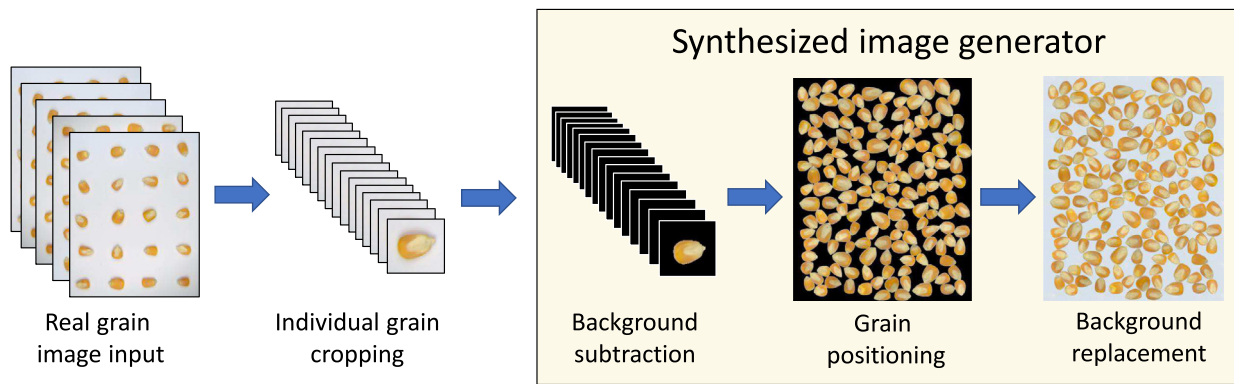
Fig. 1. Overall pipeline for synthesized clusters of grains generation. Firstly, each grain from the real-world images—i.e., grids of grains—is cropped. Then, the grain's area is extracted using background subtraction. Next, grains are inserted into an empty synthetic image by the positioning algorithm; and finally, a custom background is applied to the synthesized image.

level of freezing (mild, moderate, and severe). Two machine learning models (KNN and SVM) and one convolutional neural network model were tested using hyperspectral imaging. The results obtained better quantitative values for the convolutional network model.

Also, in Li et al. (2019), the authors propose a deep learning-based technique to discriminate different defective types of corns. In the same way, in Ni et al. (2019) the authors propose the use of very deep convolutional networks, such as VGG Simonyan and Zisserman (2015) and Residual Network (ResNet) He et al. (2016), which performs the task of classification on dual touching kernels. Recently, another classification technique has been presented in Javanmardi et al. (2021), it follows a multiclass classification approach of corn grain using a convolution network to extract the most representative characteristics. The authors just use the CNN to extract representative features, but not to perform the whole classification process.

3. Proposed approach

This section presents first the strategy proposed to generate a real-world and synthesized image dataset; then, it gives details of the Mask R-CNN network used to perform the instances segmentation. Finally, it presents the proposed corn kernels classification network used to classify each instance into one of the three categories.

3.1. Dataset generation

On the contrary, to other pattern recognition problems, where there is a large literature and available datasets (e.g., pedestrian or vehicle detection and classification), on the seed classification domain there are just a few approaches, and most of them trained and evaluated on their datasets—in most of the cases just a few annotated images. Without a common benchmark, it is impossible to do comparisons, as well as to know the limits of the state-of-the-art approaches. In the current work, a strategy for dataset generation is proposed and the obtained dataset is released to the community for further evaluation and comparisons.[3]

The dataset generation process consists first of taking different samples of corn kernels according to the three categories considered in the current work (i.e., good grains, defective grains, and impurities). Samples from each category are manually grouped to avoid further image annotation tasks. Then, a set of images from each of the three clusters are acquired with a 12MP camera orthogonal to the background plane (A4 sheet) containing the samples in a natural lighting scenario. Each image contains just the A4 sheet with the instances regularly dis-

tributed; a rectangular grid is used to make easy the split-up of each image into (5×4) cells, and each cell contains just a single grain. This process is repeated for the three categories, obtaining approximately 1000 instances per category. Finally, images from each ROI (cells in the A4 sheet) are stored as single images for further processing—see *individual grain cropping* illustration in Fig. 1.

Once a large dataset of single instances, of the three categories considered in this work, has been acquired synthesized corn kernel clusters are generated. The goal of this process is to avoid the time-consuming, and possible source of error, image annotation task. The synthesized image generator (see Fig. 1) allows the generation of corn kernel cluster images of different sizes containing different percentages of the three categories. Each instance is randomly placed and rotated avoiding overlapping between them. The user-defined parameters for the synthesized image generator are the size of the resulting image, the number of images, the minimum and the maximum number of instances, the percentage of goods, defective and impurities, the percentage of grains touching and cutting the image boundary, and background color. In case the number of instances is too large for the size of the defined image, the scale of the grains can be configured to fulfill the requirements. The synthesized image generator consists of three stages as shown in Fig. 1 (right): background subtraction, grain positioning, and background replacement.

Background subtraction is performed in the HSV color space, with thresholding parameters empirically obtained; it is robust to different lighting conditions tested during the image acquisition step. After the image thresholding, the obtained grain mask is refined with the application of morphological operators to eliminate shadows and improve contour definition. Then, the obtained grain mask is used to extract the points that define the contour of the grain; this contour is used as ground truth for training the instance segmentation algorithm. This background subtraction and contour estimation is applied over every single image obtained from the splitting-up of the original grid cells. **Grain positioning** randomly selects masks, obtained from the background subtraction step mentioned above, according to the parameters provided to the generator. Then, it inserts them into an empty image of the specified dimensions applying a random rotation, in each iteration, the algorithm finds available spaces in the synthesized image being generated and scrolls the grain until it does not overlap with previous masks. The relative position of grains contours is stored in a data structure to generate the grains annotated file—ground truth. Finally, a **background replacement** stage is applied to replace the provided background in the full synthesized image.

In addition to the acquisition process mentioned above, where seeds are placed in cells of a regular grid, a small set of images containing a cluster of corn kernels has been also acquired. Seeds and impurities in this small set of images are manually annotated, using the Labelbox crowdsourcing annotation tool, and used as ground truth to evaluate

---

[3] Code and dataset will be available through the following link: https://github.com/vision-cidis/CK-CNNLW.
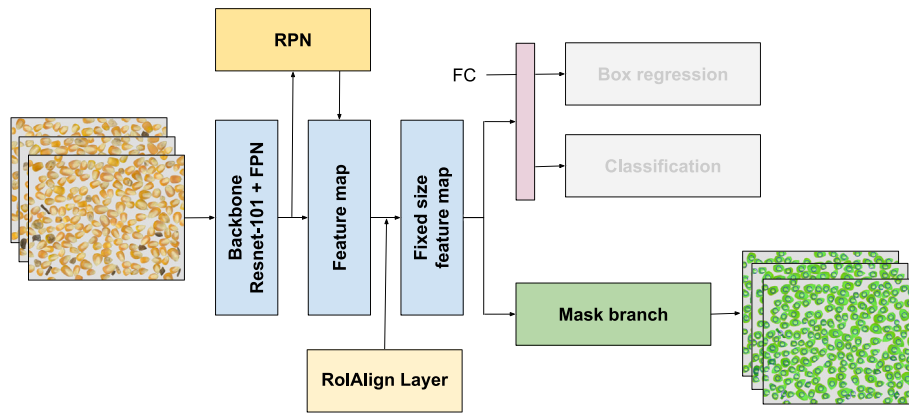
Fig. 2. Mask R-CNN architecture used for grain instance segmentation. Images of synthesized clusters of corn grains are shown as an example (classification module is not used).

the performance of the segmentation algorithm. This evaluation dataset contains 3 images with about 204 instances each.

3.2. Image segmentation

As presented in Section 2 there are different approaches in the literature to segment instances from a given image. Among the different options, the Mask R-CNN architecture He et al. (2017) has become one of the best deep-learning-based options. The Mask R-CNN framework consists of three stages (see Fig. 2). First, the backbone extracts feature maps from the input images. Second, the feature maps generated by the backbone are sent to the Region Proposal Network (RPN) to generate Regions of Interest (ROI). Third, the ROIs generated by RPN are mapped to extract the corresponding target features on the shared feature maps, and then map them to a fully connected layer, for the target classification and segmentation instances. The process generates the classification scores, bounding boxes, and segmentation masks. More details on the Mask R-CNN implementation are given in Section 4.

In general, the use of deep learning algorithms requires having large sets of annotated images for the training stage. In the particular case of the current work, where there are not a large set of annotated images available, a manual annotation process is required. It is a challenging task that requires a lot of time and resources. To avoid this problem, in our previous work Velesaca et al. (2021a) the usage of a synthesized data set has been evaluated. It contains just 20 synthesized images and has shown appealing results. In the current work, we propose to follow the same strategy, hence a large set of single instances has been acquired allowing the generation of a scale-annotated dataset. As mentioned in Section 3.1 the algorithm proposed for a synthesized cluster of grains generation allows generating datasets with different: amounts of images, percentages of instances, and seeds distribution, among other variables. It is expected that as large is the dataset as better the result; hence, to define how many images should be used for training the Mask R-CNN algorithm different datasets have been generated with an increasing number of synthesized images; more information on these datasets is given in Table 2, together with the information of other datasets (i.e., Velesaca et al. (2020), Velesaca et al. (2021a)) used for comparison in the experimental result section.

3.3. Classification

Regarding the classification stage, a novel and lightweight architecture has been designed in order to facilitate its usage in an embedded application, reducing computational complexity. The proposed architecture, referred to as CK-CNNLW, is shown in Fig. 3. The architecture consists of only four layers of convolutional operations with the corresponding activation functions. The architecture uses a leaky rectified

linear unit, which is an activation function based on a ReLU that serves to avoid sparse gradients at training time, not pruning the negative part to zero during backpropagation. To measure the loss of the training a cross-entropy loss function has been used. The model ends with two fully connected layers to flatten and connect the features extracted onto the output layer. The CK-CNNLW network has been trained with a learning rate of 0.00027, in order to generalize the model, which provides a faster convergence.

The model could be used to perform a multiclass classification problem, changing only the last layer according to the number of classes to be classified. In our case, this model has been used for a 2-class classification (good or defective corn kernels) and a 3-class classification (good corn kernels, defective corn kernels, and impurities). The CK-CNNLW network has been trained from scratch using Nesterov ADAM (NADAM) optimizer with a learning rate of 0.000275, and leaky-relu of 0.15 for average pooling and max-pooling layers. A Titan V NVIDIA GPU was used to train our model, this process takes only 5 h given the simplicity of our model. The obtained results have been compared with the previous approaches, including networks of the state of the art like VGG Simonyan and Zisserman (2015) or ResNet He et al. (2016), these comparisons are presented in the next section.

4. Experimental results

This section presents results obtained with the proposed pipeline, including both evaluations of the segmentation trained with synthesized images as well as classification of the segmented instances (i.e., good kernels, defective kernels, and impurities).

Table 2

Datasets used for training segmentation algorithm (Mask R-CNN) (Real[†]: dataset with real images Velesaca et al. (2020); Syn[†]: dataset with synthetized images Velesaca et al. (2021a); Syn$_i$: current work).

| Dataset	# of images		# of instances					
	Train	Val	Train			Val		
			Good	Def.	Imp.	Good	Def.	Imp.
Real[†]	16	4	2835	466	84	673	83	23
Syn[†]	16	4	2384	494	80	600	124	20
Syn$_1$	24	6	3543	734	120	896	185	30
Syn$_2$	32	8	4741	974	160	1189	245	40
Syn$_3$	40	10	5946	1217	200	1485	307	50
Syn$_4$	48	12	7145	1465	240	1785	367	60
Syn$_5$	56	14	8365	1715	280	2083	430	70
Syn$_6$	72	18	10,768	2210	359	2661	550	90
Syn$_7$	96	24	14,359	2949	479	3557	736	120
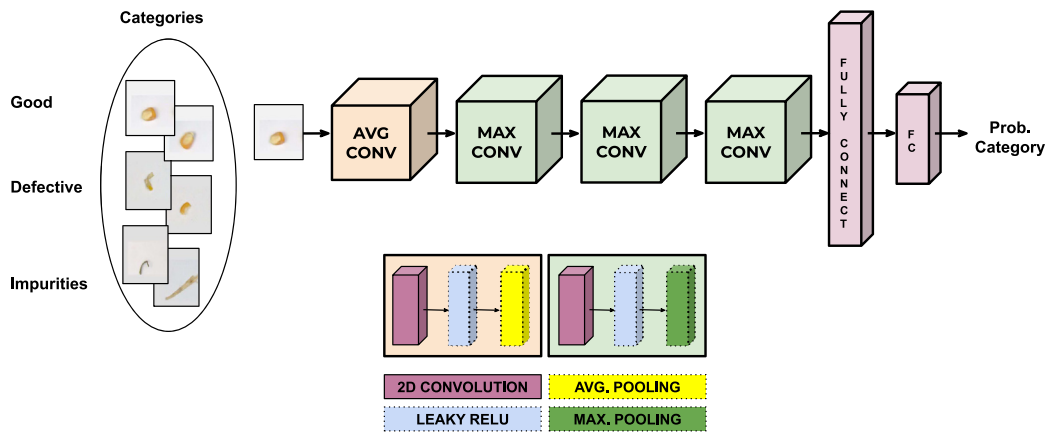
Fig. 3. Proposed lightweight classification architecture.

4.1. Instance segmentation

The results obtained by training the Mask R-CNN network He et al. (2017) using real-world and synthesized clusters of corn kernels datasets are evaluated in this section. The Mask R-CNN network implementation used in this work is based on ResNet-101 as the backbone and pre-trained COCO weight. The images in the training dataset have been resized to 1024×1024 to reduce the computational cost of the entire process. The distribution of images used in the different datasets is shown in Table 2, where the number of images for training and validation is varied, and the number of instances per image is also shown. Fig. 2 shows the architecture of the Mask R-CNN network used in this work.

The performance on the different datasets is evaluated as follows: *i*) by taking into account the number of grain instances correctly detected; *ii*) using the intersection over union (IoU), and *iii*) through the average precision (AP) in IoU 75% (AP_{75}), and the average value of IoU 50% to 95% with a step size of 5% (AP@[0.5:0.95]). Tables 3 and 4 show experimental results obtained with the Mask R-CNN network trained with the different datasets. Table 3 shows the obtained number of instances and the IoU metric computed on the set of three manually annotated images, considered as ground truth, mentioned in Section 3.1. The first row (GT) corresponds to the number of instances per ground truth image; the second row shows results obtained when trained with the dataset presented in Velesaca et al. (2020)—it should be mentioned this dataset only contains manually annotated real-world images; the third row corresponds to the results obtained from the evaluation of the Mask R-CNN network when trained with a synthesized dataset Velesaca et al. (2021a). Results from rows 4 to 10, correspond

to the results obtained with the Mask R-CNN network trained with different datasets containing synthesized images, the number of images was increased in the training stage of each dataset. On the other hand, Table 4 shows the average precision (AP) in IoU 75% (AP_{75}), and the average value of IoU 50% to 95% with a step size of 5% (AP@[0.5:0.95]).

Looking at the results depicted in Table 3, it can be appreciated that the number of detected instances is the same when real-world images or synthesized images are considered for training; on the contrary, it does not happen in the case of the IoU metric. The results of the IoU metric show better performance, in all cases when the synthesized dataset is considered, an improvement of up to almost 1.4% can be observed in the first test image. On the other hand, the AP results for this case study are shown in Table 4, where AP@[0.5:0.95] and AP_{75} metric values are shown. It can be observed that the Mask R-CNN trained with synthesized images presents a better performance in the AP@[0.5:0.95] and AP_{75} metrics. The results obtained in the IoU and AP metrics show that the use of synthesized datasets allows a high percentage of precision to correctly delimit the area and contour of the corn kernels, confirming the effectiveness and validity of the proposed approach.

In order to determine the dataset with the best configuration and performance, four metrics (i.e., number of instances, IoU, AP@[0.5:0.95], and AP_{75}) are considered, although each metric is individually evaluated, the metric of the number of instances in the present work has the highest priority, importance, and is used to discard datasets show takes in the number of detected instances directly affect the value obtained in the other three evaluated metrics. For the aforementioned, among the results obtained in the evaluation of the different datasets, the Re, Syn_4, and Syn_6 datasets show the best results about the number of detected instances and also taking into account the IoU metric, the Syn_4 dataset presents the best performance of the 3 datasets for

Table 3
Results on *testing images* (manually annotated ground truth) when the Mask R-CNN network is trained with: real images (Real[†]) Velesaca et al. (2020); and synthesized clusters dataset (Syn[†]) Velesaca et al. (2021a) —GT: Ground Truth.

Dataset	# of instances				IoU			
	T1	T2	T3	Sum	T1	T2	T3	Avg
GT	200	190	223	613	–	–	–	–
Real[†]	**199**	**189**	**215**	**603**	**0.901**	**0.897**	**0.898**	**0.899**
Syn[†]	198	188	215	601	0.914	0.911	0.900	0.908
Syn_1	199	188	215	602	0.909	0.913	0.897	0.906
Syn_2	199	189	214	602	0.909	0.910	0.897	0.906
Syn_3	198	189	215	602	0.915	0.914	0.900	0.910
Syn_4	**198**	**189**	**216**	**603**	**0.913**	**0.913**	**0.898**	**0.908**
Syn_5	199	188	215	602	0.909	0.913	0.897	0.906
Syn_6	**198**	**189**	**216**	**603**	**0.911**	**0.911**	**0.896**	**0.906**
Syn_7	198	189	215	602	0.913	0.909	0.914	0.909

Table 4
Results using the AP metric on *testing images* (manually annotated ground truth) when the Mask R-CNN network is trained with real images (Real[†]) Velesaca et al. (2020); and synthesized clusters dataset (Syn[†]) Velesaca et al. (2021a).

| Dataset	AP@[0.5:0.95]				AP_{75}			
	T1	T2	T3	Avg	T1	T2	T3	Avg
Real[†]	**0.790**	**0.800**	**0.780**	**0.793**	**0.964**	**0.978**	**0.950**	**0.958**
Syn[†]	0.830	0.830	0.790	0.818	0.980	0.984	0.945	0.964
Syn_1	0.790	0.800	0.760	0.784	0.978	0.984	0.937	0.966
Syn_2	0.790	0.820	0.770	0.793	0.979	0.973	0.932	0.961
Syn_3	0.810	0.830	0.780	0.804	0.975	0.989	0.941	0.968
Syn_4	**0.800**	**0.830**	**0.770**	**0.799**	**0.970**	**0.989**	**0.941**	**0.967**
Syn_5	0.800	0.810	0.760	0.791	0.985	0.984	0.928	0.965
Syn_6	**0.810**	**0.820**	**0.770**	**0.802**	**0.974**	**0.978**	**0.936**	**0.963**
Syn_7	0.800	0.820	0.770	0.797	0.974	0.984	0.941	0.966
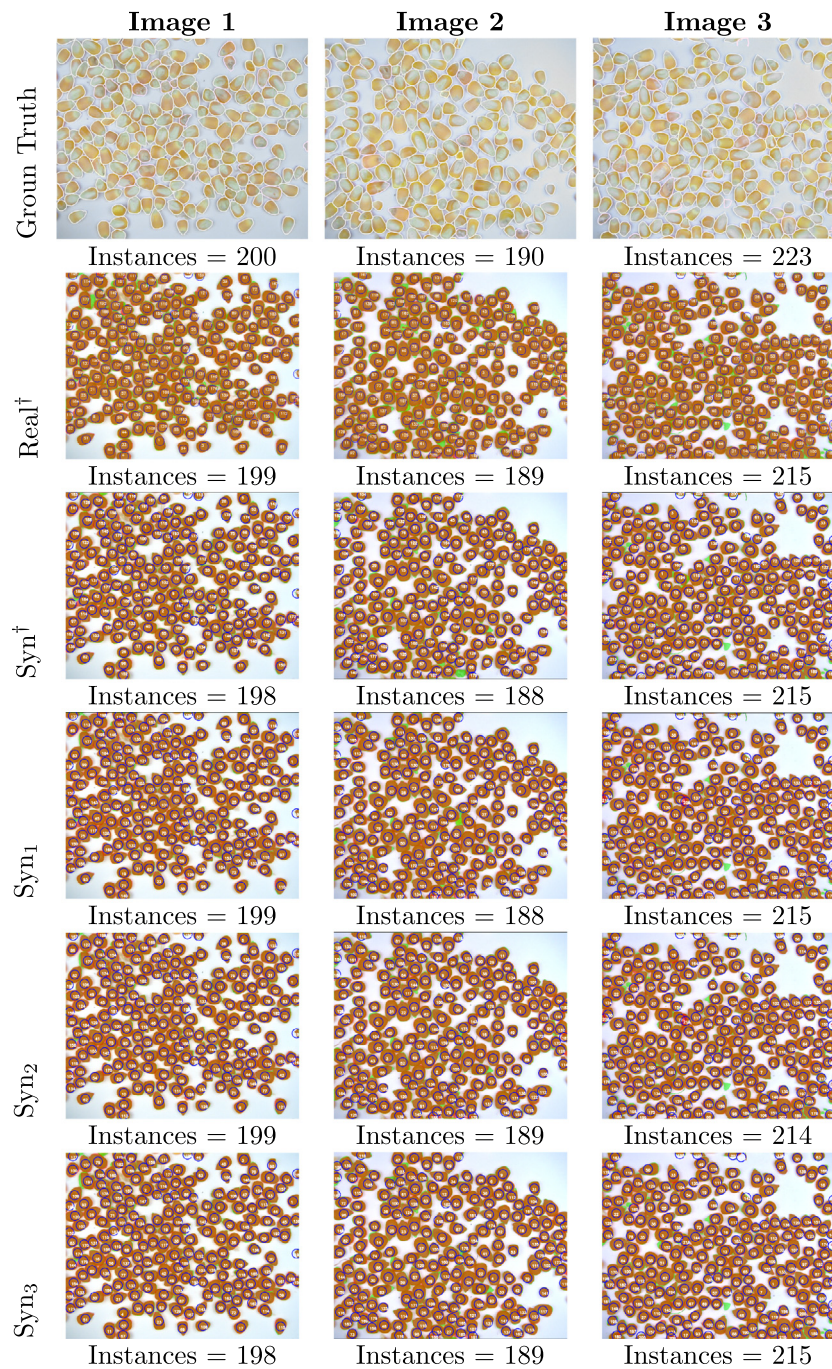
both metrics. Taking into account the previous results evaluated of the two metrics, it was observed that in the AP@[0.5:0.95] and AP_{75} metrics for the datasets Re, Syn_4, and Syn_6, it was observed that for AP@[0.5:0.95] and AP_{75} metrics the Syn_4 dataset obtains the best results on average. In general, in the four evaluated metrics, it was observed that the configuration presented by the Syn_4 dataset obtains the best performance.

Finally, the qualitative results and the annotations of the ground truth in the three test images are shown in Fig. 4, where the number of instances predicted by Mask R-CNN and the values of the ground truth are presented. For easy visualization of the qualitative evaluation, the area of each grain segmented by the Mask R-CNN is colored brown while the manual annotations are shown in green. Additionally, a blue circle has been used to highlight each individual instance along with the corresponding instance number, to verify that there are no duplicate or poorly segmented kernels.

4.2. Classification

This section presents results obtained with the proposed lightweight network for classifying corn grains according to their quality. This approach receives as an input an instance obtained in the previous instance segmentation stage. The proposed approach is evaluated in two scenarios: *i*) two-class classification problem (i.e., good and defective corn kernels), and *ii*) three-class classification problem (i.e., good, defective, and impurity). The results obtained with both approaches
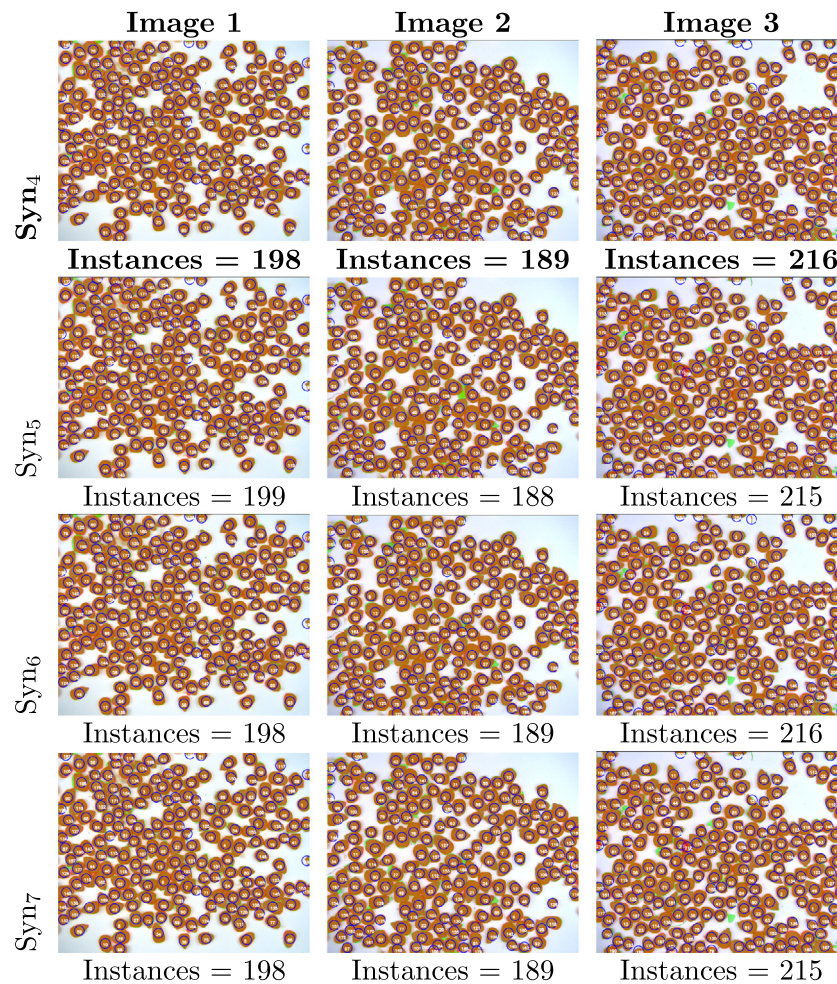
Fig. 4. Segmentation results obtained on *testing images* (manually annotated ground truth images) when Mask R-CNN is trained with real (Real[†]) and synthesized (Syn[†]) datasets— the number of obtained instances is shown in each case. (*1st row*) Ground truth labeled with Labelbox. (*2nd row*) Results when the real-world image dataset (Real[†]) Velesaca et al. (2020) is used for training. (*3rd row*) Results when the synthesized cluster dataset (Syn[†]) Velesaca et al. (2021a) is used for training. (*4th to 10th row*) Results when different synthesized cluster datasets are used for training (see Table 2).

are better than those obtained with the state-of-the-art architectures, such as VGG16, ResNet50, and Mask R-CNN, which were used for quantitative comparisons.

The two-class classification approach has been trained using a set of 3710 images (2000 images of good and defective grains for training, 1350 images of good and defective grains for testing, and 360 images of good and defective grains for validating the model after the training process). On the other hand, in the three-class classification scenario, the proposed approach has been trained using 6800 images (2100 images of good corns, 2600 images of defective corns, and 2100 images of impurities), 1800 images are used for testing the training process (600 images of good corns, 600 images of defective corns and 600

images of impurities). Finally, regarding the validation process, a total of 2060 images are considered (700 images of good corns, 700 images of defective corns, and 660 images of impurities). All the images available for the training and testing processes have been enlarged through the application of data augmentation techniques, performing operations of rotation, flipping, and horizontal and vertical shifts, to ensure the effectiveness and faster convergence of the model to be trained.

As mentioned above, the proposed model is much lighter than the corn grain classification architecture proposed by Velesaca et al. (2020), with which the extraction process of the most relevant characteristics is carried out more effectively. Results from the proposed light-weight network, CK-CNNLW, can be seen in Table 5 and Table 6, for the

Table 5
Results of classification stage for the two-class scenario.

| Network | Good | Defective | Avg. | Avg. Test Time | # of Net. |
	Corn	Corn	Acc	per Image (*sec*)	Param.
Mask R-CNN	0.962	0.644	0.803	0.0080	63,738 K
VGG16	0.950	0.917	0.933	0.0097	134,268 k
ResNet50	0.906	0.917	0.911	0.0106	23,591 K
CK-CNN	**0.956**	**0.933**	**0.945**	**0.0074**	**3306 K**
CK-CNNLW	**0.98**	**0.983**	**0.965**	**0.0063**	**548 K**

Table 6
Results of classification stage for the three-class scenario.

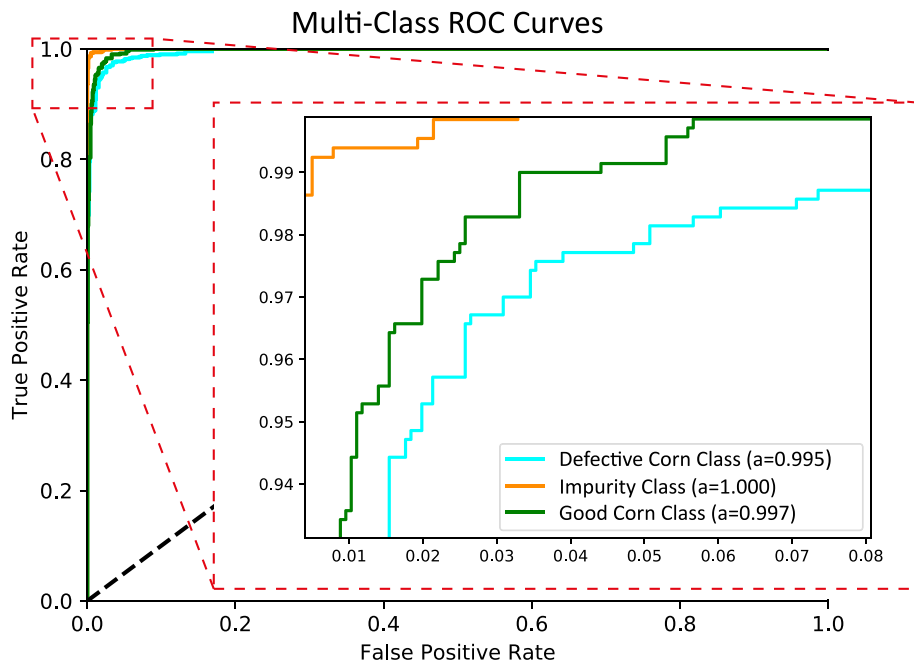| Network | Good | Def. | Imp. | Avg. | Avg. Test Time | # of Net. |
	Corn	Corn		Acc	per Image (*sec*)	Param.
Mask R-CNN	0.960	0.695	0.286	0.647	0.0140	63,738 K
VGG16	0.974	0.876	0.819	0.890	0.0112	134,272 k
ResNet50	0.986	0.860	0.931	0.925	0.0121	23,593 K
CK-CNN	**0.979**	**0.900**	**0.973**	**0.956**	**0.0095**	**3306 K**
CK-CNNLW	**0.990**	**0.930**	**0.986**	**0.967**	**0.0075**	**597 K**

Fig. 5. ROC curves of the proposed CK-CNNLW architecture for the three-class classification case.

classification problem of two and three classes respectively. These tables show the results of the proposed architecture together with the results of the state-of-the-art models, such as the Mask R-CNN He et al. (2017), VGG-16 Simonyan and Zisserman (2015), ResNet-50 He et al. (2016) and CK-CNN Velesaca et al., 2020. As can be seen, the proposed architecture shows better quantitative results than all the previous approaches. Furthermore, it should be noted that the proposed architecture requires almost six times fewer parameters than our previous approach (CK-CNN Velesaca et al. (2020)) and more than forty times fewer parameters than the rest of the architectures.

In addition to the quantitative values presented in previous tables, the ROC curves corresponding to the results obtained with the proposed architecture for the three-class classification case are depicted in Fig. 5.

Figs. 6, 7 and 8 present comparisons of ROC curves for the defective, impurity, and good corn classes between the proposed approach and our previous work (CK-CNN Velesaca et al. (2020)). It can be appreciated, in all the cases, that the proposed approach reaches a higher performance and precision when compared with CK-CNN Velesaca et al. (2020).

5. Conclusions

This paper presents a novel approach to classify clusters of touching corn kernels, which may contain good and defective kernels together with impurities; it consists first of a simple but efficient strategy to avoid the time-consuming annotation task for training the instance
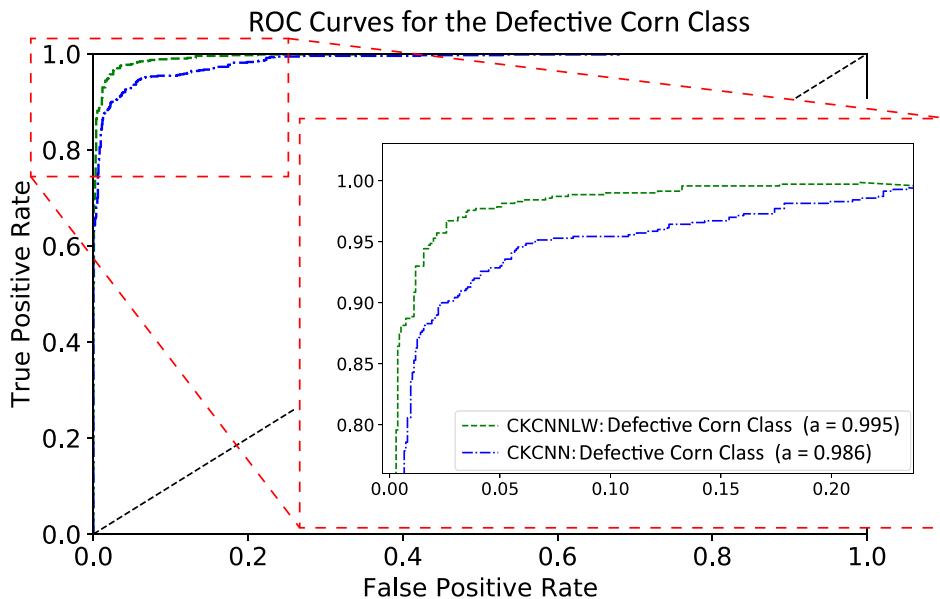

Fig. 6. ROC curves of the CK-CNN and the proposed CK-CNNLW model for the defective corn class (three-class classification case).
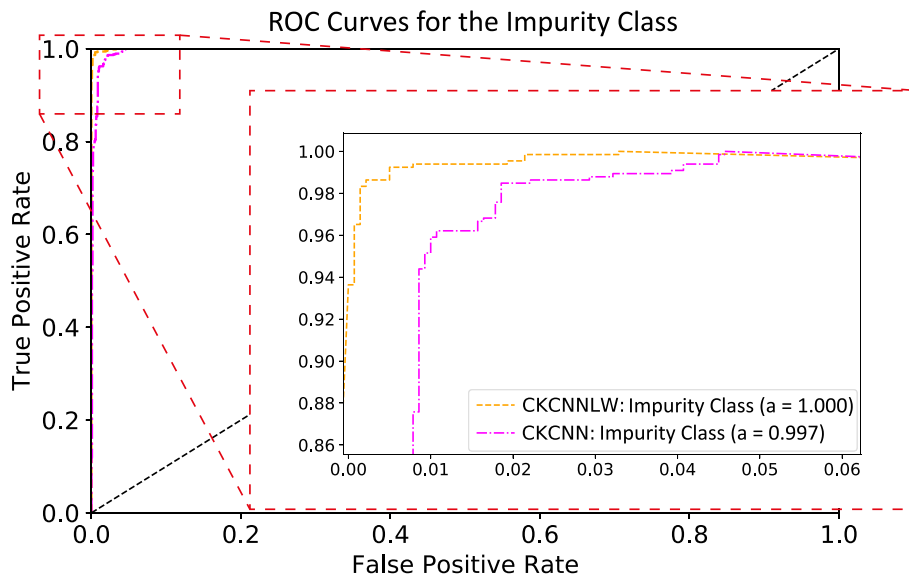
Fig. 7. ROC curves of the CK-CNN and the proposed CK-CNNLW model for the impurity class (three-class classification case).
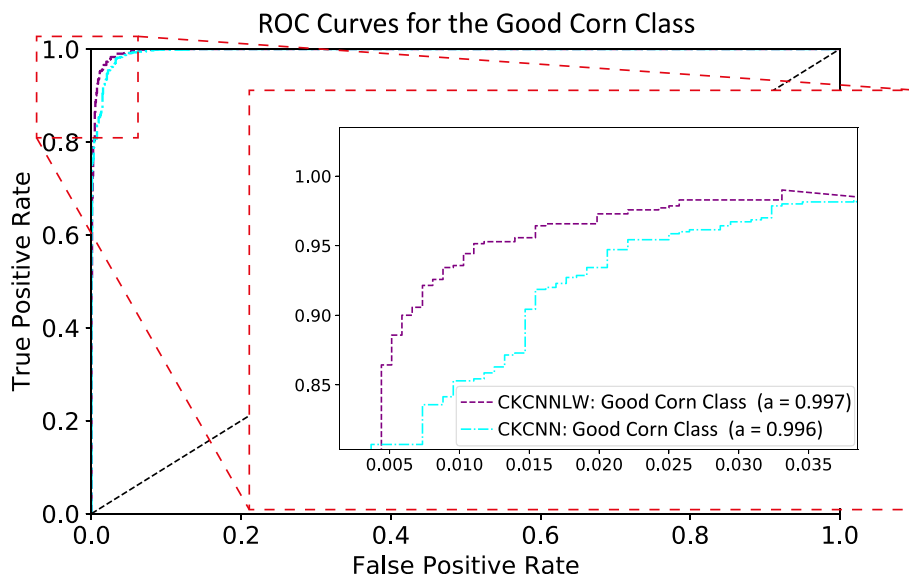

Fig. 8. ROC curves of the CK-CNN and the proposed CK-CNNLW model for the good corn class (three-class classification case).

segmentation Mask R-CNN algorithm. Results obtained by the Mask R-CNN network trained with synthetized datasets reach better performance, both IoU and AP than when training with manually annotated images. Then, as a *sec*ond contribution, a novel and lightweight network is proposed (CK-CNNLW) to classify the obtained instances. This network is considerably lighter than previous approaches (six times fewer parameters than our previous approach and more than forty times fewer parameters than other state-of-the-art architectures). Furthermore, obtained results show better performance than all previous approaches. In future work we will try to develop our own segmentation model, trying to obtain a lightweight architecture. Additionally, we plan to extend the proposed solution to other types of grains.

CRediT authorship contribution statement

Patricia L. Suárez: Software, Methodology, Validation, Writing – original draft, Investigation, Visualization, Writing – review & editing.

Henry O. Velesaca: Investigation, Software, Validation, Writing – original draft. **Dario Carpio:** Software, Writing – original draft. **Angel D. Sappa:** Conceptualization, Resources, Supervision, Investigation, 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.

References

Adke, S., Haro Von Mogel, K., Jiang, Y., Li, C., et al., 2020. Instance segmentation to estimate consumption of corn ears by wild animals for gmo preference tests. Front. Artif. Intelligen. 3, 1–14.

Ali, A., Qadri, S., Mashwani, W.K., Brahim Belhaouari, S., Naeem, S., Rafique, S., Jamal, F., Chesneau, C., Anam, S., 2020. Machine learning approach for the classification of corn seed using hybrid features. Int. J. Food Prop. 23, 1110–1124.

Ambrose, A., Kandpal, L.M., Kim, M.S., Lee, W.-H., Cho, B.-K., 2016a. High speed measurement of corn seed viability using hyperspectral imaging. Infrared Phys. Technol. 75, 173–179.

Ambrose, A., Lohumi, S., Lee, W.-H., Cho, B.K., 2016b. Comparative nondestructive measurement of corn seed viability using fourier transform near-infrared (ft-nir) and raman spectroscopy. Sensors Actuators B Chem. 224, 500–506.

Arboleda, E.R., Fajardo, A.C., Medina, R.P., 2018. An image processing technique for coffee black beans identification. IEEE Int. Conf. On innovative Research and Development, pp. 1–5.

Aukkapinyo, K., Sawangwong, S., Pooyoi, P., Kusakunniran, W., 2019. Localization and classification of rice-grain images using region proposals-based convolutional neural network. Int. J. Autom. Comput. 1–14.

Bacon, C., 2005. Confronting the coffee crisis: can fair trade, organic, and specialty coffees reduce small-scale farmer vulnerability in northern Nicaragua? World Dev. 33, 497–511.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Cheng, X., Vella, A., Stasiewicz, M.J., 2019. Classification of aflatoxin contaminated single corn kernels by ultraviolet to near infrared spectroscopy. Food Control 98, 253–261.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20, 273–297.

Cover, T., Hart, P., 1967. Nearest neighbor pattern classification. IEEE Trans. Inf. Theory 13, 21–27.

Daskalov, P., Kirilova, E., Georgieva, T., 2018. Performance of an automatic inspection system for classification of fusarium moniliforme damaged corn seeds by image analysis. MATEC Web of Conferences. 210. EDP Sciences, p. 02014.

Effendi, M., Jannah, M., Effendi, U., 2019. Corn quality identification using image processing with k-nearest neighbor classifier based on color and texture features. IOP Conference Series: Earth and Environmental Science, pp. 12–66.

Friedman, J., Hastie, T., Tibshirani, R., 2000. Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). Ann. Stat. 28, 337–407.

Ganesh, P., Volle, K., Burks, T., Mehta, S., 2019. Deep orange: mask R-CNN based orange detection and segmentation. IFAC-PapersOnLine 52, 70–75.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778.

He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask R-CNN. IEEE Int. Conf. On Computer Vision, pp. 2961–2969.

Huang, S., Fan, X., Sun, L., Shen, Y., Suo, X., 2019. Research on classification method of maize seed defect based on machine vision. J. Sens. 2, 1–10.

Javanmardi, S., Ashtiani, S.-H.M., Verbeek, F., Martynenko, A.I., 2021. Computer-vision classification of corn seed varieties using deep convolutional neural network. J. Stored Prod. Res. 92.

Kar, A., Kulshreshtha, P., Agrawal, A., Palakkal, S., Boregowda, L.R., 2019. Annotation-free quality estimation of food grains using deep neural network. 30th British Machine Vision Conference, pp. 1–12.

Kaur, S., Singh, D., 2015. Geometric feature extraction of selected rice grains using image processing techniques. Int. J. Comput. Appl. 124, 41–46.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In: Pereira, F., Burges, C., Bottou, L., Weinberger, K. (Eds.), Advances in Neural Information Processing Systems. 25. Curran Associates, Inc.

Li, X., Dai, B., Sun, H., Li, W., 2019. Corn classification system based on computer vision. Symmetry 11, 1–12.

Naylor, L., 2017. Auditing the subjects of fair trade: coffee, development, and surveillance in highland Chiapas. Environ. Plann. D: Soc. Spac. 35, 816–835.

Ni, C., Wang, D., Holmes, M., Vinson, R., Tao, Y., 2018. Convolution neural network based automatic corn kernel qualification. ASABE Annual Int. Meeting. American Society of Agricultural and Biological Engineers.

Ni, C., Wang, D., Vinson, R., Holmes, M., Tao, Y., 2019. Automatic inspection machine for maize kernels based on deep convolutional neural networks. Biosyst. Eng. 178, 131–144.

Paliwal, J., Visen, N., Jayas, D., White, N., 2003. Cereal grain and dockage identification using machine vision. Biosyst. Eng. 85, 51–57.

Paulsen, M.R., Singh, M., Singh, V., 2019. Measurement and maintenance of corn quality. Corn. Elsevier, pp. 165–211.

Pearl, J., 1985. Bayesian netwcrks: a model cf self-activated memory for evidential reasoning. Proceedings of the 7th Conference of the Cognitive Science Society. University of California, Irvine, CA, USA, pp. 15–17.

Qiu, G., Lü, E., Lu, H., Xu, S., Zeng, F., Shui, Q., 2018. Single-kernel ft-nir spectroscopy for detecting supersweet corn (zea mays l. saccharata sturt) seed viability with multivariate data analysis. Sensors 18, 1010.

Qiu, G., Lü, E., Wang, N., Lu, H., Wang, F., Zeng, F., 2019. Cultivar classification of single sweet corn seed using fourier transform near-infrared spectroscopy combined with discriminant analysis. Appl. Sci. 9, 1530.

Ren, S., He, K., Girshick, R., Sun, J., 2016. Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. 39, 1137–1149.

Ribeiro, S.S., Yao, J., 2019. Toward a three-way image classification model: a case study on corn grain images. IEEE Int. Symposium on Multimedia, pp. 177–1776.

Roerdink, J.B., Meijster, A., 2000. The watershed transform: definitions, algorithms and parallelization strategies. Fundamenta Informat. 41, 187–228.

Shamim, M.I., Pal, B., Arora, A.S., Pial, M.A., 2020. A deep convolutional neural network approach to rice grain purity analysis. Progress in Computing, Analytics and Networking. Springer, pp. 179–189.

Shao, H., Tang, R., Lei, Y., Mu, J., Guan, Y., Xiang, Y., 2021. Rice ear counting based on image segmentation and establishment of a dataset. Plants 10, 1–12.

Shrestha, B.L., Kang, Y.-M., Yu, D., Baik, O.-D., 2016. A two-camera machine vision approach to separating and identifying laboratory sprouted wheat kernels. Biosyst. Eng. 147, 265–273.

Simonyan, K., Zisserman, A., 2015. Very deep convolutional networks for large-scale image recognition. 3rd International Conference on Learning Representations. ICLR, San Diego, CA, USA.

Singh, K.R., Chaudhury, S., 2020. A cascade network for the classification of rice grain based on single rice kernel. Complex & Intelligent Systems, pp. 1–14.

Toda, Y., Okura, F., Ito, J., Okada, S., Kinoshita, T., Tsuji, H., Saisho, D., 2020. Training instance segmentation neural network with synthetic datasets for crop seed phenotyping. Communicat. Biol. 3, 1–12.

Velesaca, H.O., Mira, R., Suarez, P.L., Larrea, C.X., Sappa, A.D., 2020. Deep learning based corn kernel classification. IEEE Int. Conf. on Computer Vision and Pattern Recognition Workshops, pp. 294–302.

Velesaca, H.O., Carpio, D., Suarez, P.L., Sappa, A.D., 2021a. Synthesized image datasets: towards an annotation-free instance segmentation strategy. 16th Int. Symposium on Visual Computing, pp. 1–13.

Velesaca, H.O., Suárez, P.L., Mira, R., Sappa, A.D., 2021b. Computer vision based food grain classification: a comprehensive survey. Comput. Electron. Agric. 187, 106287. https://doi.org/10.1016/j.compag.2021.106287.

Williams, P.J., Kucheryavskiy, S., 2016. Classification of maize kernels using nir hyperspectral imaging. Food Chem. 209, 131–138.

Wilson, B.R., 2010. Indebted to fair trade? Coffee and crisis in Nicaragua. Geoforum 41, 84–92.

Wold, S., Sjöström, M., Eriksson, L., 2001. Pls-regression: a basic tool of chemometrics. Chemom. Intell. Lab. Syst. 58, 109–130.

Wu, A., Zhu, J., Yang, Y., Liu, X., Wang, X., Wang, L., Zhang, H., Chen, J., 2018. Classification of corn kernels grades using image analysis and support vector machine. Adv. Mech. Eng. 10, 1–9.

Yang, X., Hong, H., You, Z., Cheng, F., 2015. Spectral and image integrated analysis of hyperspectral data for waxy corn seed variety classification. Sensors 15, 15578–15594.

Yuan, Y., Wang, W., Chu, X., Xi, M., 2016. Selection of characteristic wavelengths using spa and qualitative discrimination of mildew degree of corn kernels based on svm. Spectrosc. Spectr. Anal. 36, 226–230.

Zhang, J., Dai, L., Cheng, F., 2021a. Corn seed variety classification based on hyperspectral reflectance imaging and deep convolutional neural network. J. Food Measure. Characterizat. 15, 484–494.

Zhang, J., Dai, L., Cheng, F., 2021b. Identification of corn seeds with different freezing damage degree based on hyperspectral reflectance imaging and deep learning method. Food Anal. Methods 14, 389–400.

Zhu, F., Yao, H., Hruska, Z., Kincaid, R., Brown, R., Bhatnagar, D., Cleveland, T., 2015. Classification of corn kernels contaminated with aflatoxins using fluorescence and reflectance hyperspectral images analysis. Sensing for Agriculture and Food Quality and Safety VII. SPIE, p. 9488.

Zhu, H., Yue, J., Li, Z., Zhang, Z., 2018. Multi-kernel dictionary learning for classifying maize varieties. Int. J. Agricult. Biol. Eng. 11, 183–189.