This CVPR 2020 workshop paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

Deep Learning based Corn Kernel Classification

Henry O. Velesaca*[1], Raúl Mira*[1], Patricia L. Suárez*[1], Christian X. Larrea[1], Angel D. Sappa[1,2]

[1]Escuela Superior Politécnica del Litoral, ESPOL,
Facultad de Ingeniería en Electricidad y Computación, CIDIS,
Campus Gustavo Galindo, 09-01-5863, Guayaquil, Ecuador

[2]Computer Vision Center, Edifici O, Campus UAB,
08193, Bellaterra, Barcelona, Spain

{hvelesac,raualmir,plsuarez,clarrea}@espol.edu.ec, sappa@ieee.org

Abstract

This paper presents a full pipeline to classify sample sets of corn kernels. The proposed approach follows a segmentation-classification scheme. The image segmentation is performed through a well known deep learning-based approach, the Mask R-CNN architecture, while the classification is performed through a novel-lightweight network specially designed for this task—good corn kernel, defective corn kernel and impurity categories are considered. As a second contribution, a carefully annotated multi-touching corn kernel dataset has been generated. This dataset has been used for training the segmentation and the classification modules. Quantitative evaluations have been performed and comparisons with other approaches are provided showing improvements with the proposed pipeline.

1. Introduction

Cereal production, both for human and animal consumption, is one of the bases of the food pyramid industry. The agriculture industry is the primary sector in the economy of several countries, in some cases representing near 10% of their Gross Domestic Product (GDP). One of the most important and demanded grains for both humans and livestock nutrition and the raw material for agribusiness is maize. It has the largest production of all cereals all over the world[1]. To reach the highest quality standard, according to the worldwide commercial protocol that establishes the type and quality of grains, recently some approaches have been proposed trying to do the kernel inspection rigorously and automatically (e.g., [12], [15], [3]).

In the particular case of automatic corn kernel assessment for the post-harvest process, there is also a large interest from the research community. In general, corn kernel assessment evaluates constituent features (e.g., moisture, crude protein, fiber, etc.) as well as visual features such as impurity, shape (including perimeter, area, elongation, among others), color, etc. Constituent measurements are obtained using tools and machines especially devoted to such tasks, while visual features are manually extracted employing trained operators. This manual process is a time-consuming operation and cannot ensure consistency due to the difference in operator's evaluation ability [14] [8].

Before evaluating the shape or color features that characterize a given sample set, elements in that sample set should be classified into some of the following categories: defective kernels (including broken or rotten kernels), impurities and good kernels. As mentioned above this classification is generally performed by trained operators, who have to inspect samples of about 200g of kernels and then getting a table of percentages of impurities, rotten and defective kernels. This time-consuming task is prone to subjectivity of the inspectors.

Recently, with the increase in computing power that allows processing a large amount of information in a short time, some computer vision solutions have been proposed in the literature (e.g., [13], [11]). In [13] the authors present a computer vision-based corn classification system. Interesting results are obtained while corn kernels are not touching each other. The authors also present an ad-hoc approach to segment groups of touching kernels. Different *off-the-shelf* CNN based classification models are evaluated (e.g., ResNet [7], VGG [21] and AlexNet [10]), concluding the

[1]http://www.fao.org/in-action/inpho/crop-compendium/cereals-grains/en

*Authors contributed equally.

best option for their dataset is ResNet. The main drawback of this approach lies in the way corn kernels should be placed; ideally, they should not be touching or in case of touching a maximum of dual-touching kernels are allowed, leading to the need of a system for separating them. The same drawback is present in the solution proposed in [11]; in this approach, although focusing on the corn classification problem, the authors do an effort to segment groups of touching corns by proposing a novel profile segmentation algorithm. Also based on machine learning approaches, some innovative products are already on the market, such as the product offered by Cgrain[2], although an accurate solution is obtained, it is also a quite time-consuming solution since every single element (grain or impurity) should be independently inspected.

Finally, trying to speed up this pre-processing stage, as well as to reach more robust solutions able to tackle challenging sample set distributions, solutions based on the usage of multispectral and hyperspectral imaging have been proposed. For instance, in [23] hyperspectral images, covering a range from 400 nm to 1000 nm, were considered to classify different maize varieties; actually, the authors show the proposed approach can be used effectively for seed identification and classification. Trying to find the best waveband for sample set classification in [22] the authors propose a combination of five wavebands in the range from 400nm to 900nm. Finally, another hyperspectral imaging-based approach, but in this case spanning the near infrared (NIR) spectral band, has been presented in [20]. NIR hyperspectral imaging helps to identify a variety of cereal properties, which could replace conventional chemical microbial or physical tests, with a single and automated image acquisition.

In the current work, a novel computer vision-based approach is proposed for corn kernel detection and classification; it works just using images from the visible spectrum. The proposed pipeline consists of firstly segmenting elements from the given sample set into single entities that later on are classified by a lightweight network. The whole approach is robust to the distribution of elements in the given sample set; in other words, close objects of various kinds (defective kernels, impurities and good kernels) do not affect the final result. Object segmentation is based on the usage of Mask R-CNN [6], while classification is performed by a novel and compact CNN based architecture CK-CNN. The main contributions in the paper are summarized as follow:

- A dataset with both, carefully corn kernel's contour annotations and four individual category labeling (defective kernels, including broken or rotten kernels, impurities and good kernels), has been generated and released to the community—CORN-KERNEL[3].
- A lightweight CNN architecture for corn kernel classification is designed, referred to as CK-CNN. The model is trained from scratch, without pre-trained weights.

The manuscript is organized as follows. Section 2 presents related works to the segmentation and classification problems, which serve as the basis for the approach of the main modules of the proposed pipeline. Section 3 presents the approach proposed for detecting and classifying elements (defective kernels, impurities and good kernels) 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 mentioned above, in the current work the corn kernel detection and classification problem is tackled following the classic detection-classification pipeline. Hence, this section reviews the most relevant works on these topics highlighting the main characteristics of state-of-the-art approaches. Firstly, the state of the art segmentation techniques, generally used in the detection phase, are reviewed; secondly, classification approaches including both, classical and recent deep learning techniques, are summarized.

2.1. Segmentation techniques

Although image segmentation is an old and well-studied problem in the computer vision literature, the cereal kernel segmentation, in a general scenario, is a challenging and open problem. Recent publications, some of them for the corn kernel problem, have proposed *ad-hoc* solutions that could reach acceptable result in some cases (e.g., [11], [13], [3]). Trying to develop a robust and general solution, that work in unconstrained scenarios, no matter the number of corns kernel touching, the state of the art techniques are reviewed and evaluated to propose a possible solution.

Firstly, the watershed transform has been considered for the cereal kernel segmentation; it is a traditional segmentation approach widely used. The main idea of this technique comes from geography, where a grayscale image could be seen as a topographic surface in which the high-intensity values are the peaks (local maximum) and the lowest values are the valleys (local minimum). At the beginning of the process, each valley can be filled in with a different color. Then, the algorithm continues filling in valleys till regions start touching and the region's boundary is defined; this watershed process continues till the highest peaks are reached;

[2] https://www.cgrain.se

[3] The dataset is available at http://www.cidis.espol.edu.ec/es/dataset

as a result, all regions in the given image are segmented [18] [16]. Despite all the features offered by this algorithm, there are several problems when applying it to real-world images. For instance, results would depend on the variation of intensity levels in the given image, as well as on the texture and morphology of the objects to be segmented; these conditions generate an over-segmentation of the regions [16]. In the particular case of corn kernel segmentation, the kernel touching problem is a challenging situation that most of the time cannot be solved with this approach.

With the success of CNN, principally because of its result in [10], different approaches have been proposed to tackle problems that are difficult to solve by traditional computer vision techniques. One of these areas is image segmentation where different architectures have been proposed improving the state of the art results. Among the different proposals, the U-Net model [19] has been introduced for the semantic segmentation of biomedical images. The main characteristic of this architecture is that it needs very few annotated images for training, in the seminal work approximately 30 images have been considered. Another architecture proposed for instance segmentation has been presented in Mask R-CNN [6]. It is a powerful tool ([9], [25]) that follows the philosophy of Fast R-CNN and improves some characteristics by adding a new branch at the end of the model, in this way the task of segmentation, location and detection of objects work in parallel. In general terms, Mask R-CNN consists of three phases, first, the backbone network (ResNet-50 or ResNet-101) extracts the feature maps of the input images, then these maps are sent to the Region Proposal Network (RPN), where the ROIs are generated. In the third phase, these ROIs are mapped to extract the corresponding target features, which are sent to the Fully Connected Layers (FC) and the mask branch, where the classification and instance segmentation are performed respectively. The process described above generates classification scores, bounding boxes and segmentation masks [25] [6].

2.2. Classification techniques

In recent years, several machine vision-based grain classification approaches have been proposed in the literature. For instance, in [2] the authors propose a pattern recognition based technique to extract the external characteristics of kernels, whether they be the shape, color or geometry from the given images; the main limitation lies on the fact that kernels should not touch each other [2]. The kernels are classified using a backward propagation neural network (BPNN) to identify the class to which they belong.

In [17] a three-way image classification model, based on HSV color space, is presented. The approach extracts textual local binary pattern (LBP) information, for the classification of healthy or rotten grains. Also, to improve the

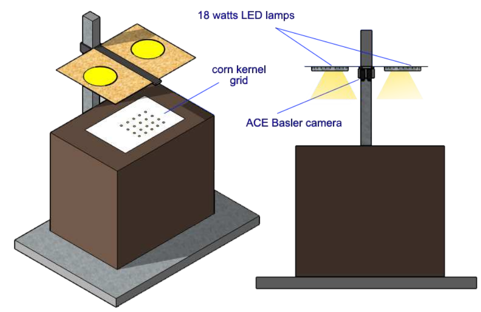

Figure 1. Illustration of the data acquisition system.

recognition rate of maize varieties, [26] proposes a multi-kernel maize varieties classification extracting the characteristics of maize grain to distinguish maize varieties. Similarly, Effendi et al. [4] present a corn quality identification system based on color and texture features to identify the quality of corns.

On the contrary to previous approaches, in [11], the authors propose a deep learning-based technique to discriminate different defective types of corns; firstly, a segmentation method is used to separate a group of touching corns. Then, 12 color features and 5 shape features are extracted for each corn object. Finally, a maximum likelihood estimator is trained to classify normal and defective corns. In the same way, in [13] the authors propose the use of very deep convolutional networks, such as VGG [21] and Residual Network (ResNet) [7], which outperforms the task of classification on dual touching kernels. The given images are firstly segmented using image processing techniques and then each element resulting from the segmentation is classified as a good or defective kernel.

VGG [21] is a ConvNet model with a very deep convolutional architecture (16-19 weight layers) and very small convolutional filters (3X3) for large scale image classification. This model has been designed for well-performing on large dataset images. The objective was to prove that deeper networks overcome state-of-art accuracy in this kind of visual representations. VGG has been originally trained on the ImageNet challenge dataset but demonstrated to do a good matching and generalization to a wide range of tasks and datasets. On the other hand, ResNet [7], is a residual learning framework that is substantially deeper than previously mentioned (8x times deeper than VGG net) with an architecture up to 152 layers; same as VGG it was trained for ISLRVC 2015 classification task and got the first place. The main characteristic of ResNet is the introduction of a so-called "identity shortcut connection" that skips one or more layers (identity maps) to fit a residual mapping instead of letting them directly fit the desired underlying mapping. The resulting architecture performs better.

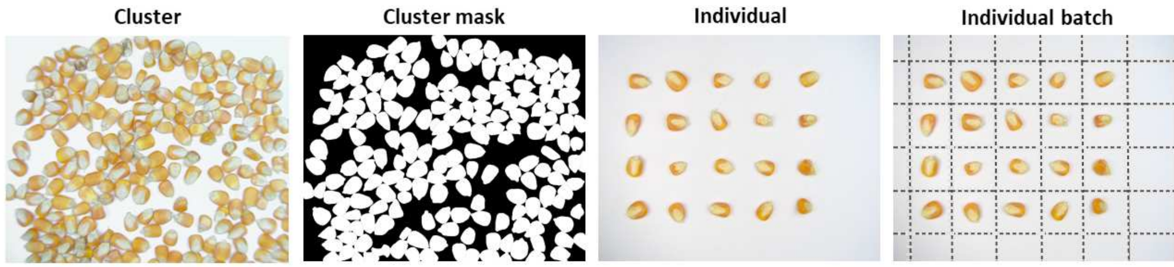

Figure 2. (*1st col*) Example of a corn kernel cluster image. (*2nd col*) Mask from the annotated contours used as a ground truth. (*3rd col*) Example of individual corn kernels. (*4th col*) Individual corn kernels split up into grid cells.

3. Proposed Approach

The proposed approach consists of two stages, firstly the elements in the given image are segmented and then they are classified into some of the following categories: good corn, defective corn (including rotten and broken corns) and impurities. Since both stages are based on learning approaches, the success of the whole pipeline would depend on the quality of the ground truth, both corn kernel's contour annotation and labeled images. Before presenting the proposed approach the dataset generated in the current work is detailed.

3.1. Dataset Generation

This section details the process carried out for data acquisition, corn kernel's contour annotation and image labeling. Corn kernels have been acquired in a controlled environment using a visible spectrum camera (Basler ACE acA645-100gc), with a resolution of 1280×1024 pixels, together with two LED lamps of 18 watts each, placed on top. The camera has been placed orthogonal to the plane containing the corn kernels; a white cardboard has been used to define a background standard for further applications. Figure 1 shows an illustration of the acquisition system. With this acquisition system, two sets of images have been recorded: i) clusters of corn kernels and ii) single corn kernels. The first set contains images of clusters of corn kernels to be used during the segmentation stage (see Fig. 2 *(1st and 2nd columns)*). The second set consists of images of a regular grid containing a single corn kernel per cell (see Fig. 2 *(3rd and 4th columns)*). With this setup a total of 523 images have been acquired; some of them have been used for corn kernel's contour annotation and the remaining for image annotation. Table 1 presents a summary of the acquired images and their usage.

Once the two datasets have been generated, the contours of every element (corn kernels and impurity) in the first dataset are carefully annotated using a crowdsourcing tool (Labelbox[4]). These carefully annotated contours are used as ground truths for training and validating the segmentation stage. Figure 2 *(2nd col)* shows the binary masks, corresponding to the contours, obtained for each element from the Labelbox annotation.

[4]https://labelbox.com/

Type	Used for	Category	Images
Cluster	Segmentation	Cluster	23
Individual	Classification	Good corn	100
Individual	Classification	Defective corn	120
Individual	Classification	Impurity	60
Total			303

Table 1. Distribution of the generated dataset. Defective corn includes rotten and broken corns.

3.2. Image Segmentation

After the datasets have been generated, the next step is to develop an approachable to segment the elements in the given image for a further classification. For this task, the first dataset (cluster of corn kernels) is considered and a deep learning-based approach, the Mask R-CNN network [6], is trained. The training process is performed by using 16 images from the corn kernel's boundary annotated set; 4 images have been used for validation and the remaining 3 images for testing (note that on average each image contains about 200 elements, mainly corn kernels). Table 2 presents a detailed description of the cluster set used for training the image segmentation algorithm. A data augmentation process, consisting of horizontal and vertical flips, together with 90-degree rotations and Gaussian blur has been considered to increase the number of images for the training process.

The code from [1] of the Mask R-CNN network [6] has been used to perform the image segmentation. This architecture generates bounding boxes and segmentation masks for each instance of the corn kernel and impurity present in the given image. This implementation is based on ResNet-101 as a backbone and pre-trained COCO weight; to retrain this network images from the corn kernel data set have been resized up to 512×512, to reduce the computational cost of

Dataset	Images	Good kernel instances	Defective kernel instances	Impurity instances
Training	16	2835	466	84
Validation	4	673	83	23
Testing	3	547	59	7
Total	23	4055	608	114

Table 2. Dataset distribution in segmentation stage.

the training process. Figure 3 shows the Mask R-CNN architecture used for corn kernel instance segmentation (note that classification module is not considered, more details are given in Section 4).

3.3. Classification

Regarding the classification, in the current work, a novel-lightweight architecture is proposed to classify a given element into some of the following classes: good corn, defective corn (including broken and rotten corns) and impurity. The usage of deep networks, such as VGG [21] or ResNet [7], as previous work (see Section 2.2), has been considered. However, after evaluating the obtained results (comparisons are presented in the next section) the design of a more specific architecture has been considered. The proposed architecture (see Fig. 4), referred to as CK-CNN, receives as an input a single element from the segmentation algorithm and consists of five layers: three convolutional layers defined with a 3×3 size kernels and two fully connected layers. The model uses a cross-entropy loss function to measure the performance of the classification model. Also, the model includes a RELU activation function after each convolution and a max pool layer to summarize the results of the convolution operation. The last two layers are fully connected, the first one receives the output of the last convolutional layer, which allows all the outputs of the convolution operation to be connected, as was done in the multilayer perceptron (MLP) technique. The last fully-connected layer enables the class score using the softmax activation function, to obtain the probability distribution that corresponds to each class type.

The proposed model supports the n-class classification problem. In our case, this model has been used for a 2-class classification (good or defective corn kernels) and for a 3-class classification (good corn kernels, defective corn kernels and impurity). The CK-CNN network has been trained from scratch using Nesterov ADAM (NADAM) optimizer with a learning rate of 0.0002, which provides a faster convergence and generalization of the model.

4. Experimental Results

This section presents the experimental results obtained with the proposed pipeline. Evaluations on the two stages: segmentation and classification are provided showing the performance of each one of them in comparison with the state of the art approaches.

4.1. Image Segmentation

As mentioned above, the image segmentation has been performed using the Mask R-CNN network, which has been trained using 16 images from the cluster of corn kernel set. The contour of each corn kernel in the image has been carefully annotated and used for training the Mask R-CNN. Table 2 shows the distribution of the dataset used in the training, validation and testing phases; it can be also appreciated the number of instances (corn kernels and impurities) present on each set. As mentioned in Section 3.2 a data augmentation process (i.e., horizontal flips, vertical flips and 90-degree rotations and Gaussian blur) has been applied to increase the number of images for the training.

After Mask R-CNN has been trained, a set of 3 images has been used for validating the obtained results as well as to compare with other approaches. Figure 5 shows the ground truth for these three images together with the results from the Mask R-CNN approach used in the current work. Additionally, results from watershed and U-Net are provided; these two approaches have been selected based on the state-of-the-art corn kernel segmentation. The output of U-Net are edges that need to be post-processed to extract single elements. This post-processing has been performed using also watershed. In all the cases brown areas correspond to overlap between regions predicted by a given algorithm and ground truth; green color is used to highlight missed regions while red regions correspond to wrong detected areas. More details on these two comparisons are provided below.

Regarding the two implementations used to compare the results from Mask R-CNN, firstly, the classical watershed algorithm has been considered. Results are presented in Fig. 5 (*2nd col*). The algorithm did not generate the expected results, several corn kernels have more than one local maximum; this fact corresponds to the kernel morphology and the light conditions in which the images have been acquired. In other words, the algorithm generates regions that do not exist and a large number of corn kernels are missed, almost half of them in the illustration of the first row.

The U-Net architecture [19] has been the second approach used to compare the results obtained with Mask R-CNN. The framework presented in [24] has been considered. It consists of a ResNet-34 as the backbone and a pre-trained ImageNet. The framework has been trained with the training set of the cluster of corn kernel set presented above. To speed up the training process, the images were resized to 512×512. Since the result of this network is a binary mask and it is necessary to determine the instances of each element from the given image watershed post-processing has

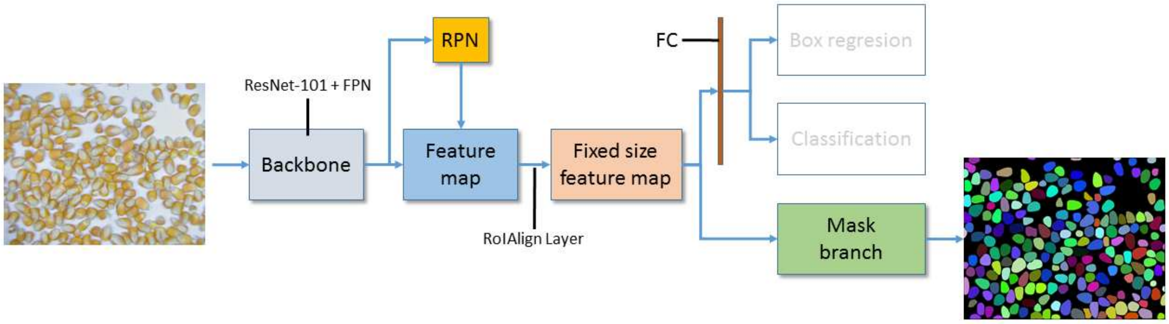

Figure 3. Mask R-CNN architecture [6] used for corn kernel instance segmentation (classification module is not used).

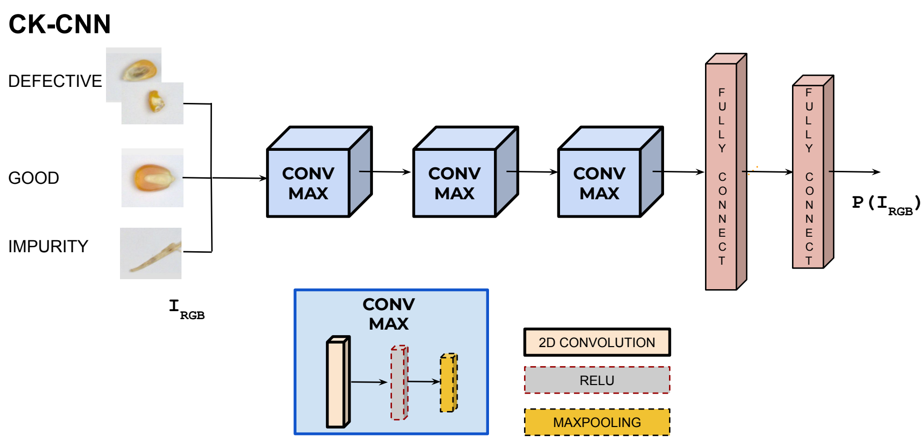

Figure 4. Proposed corn kernel classification network (CK-CNN).

been performed. This post-processing stage uses as an input the binary mask from U-Net and determines each element in the image. Figure 5 (*3rd col*) depicts results obtained from this two-stage approach. Although better results are obtained, they are not as good as the ones obtained with Mask R-CNN (see Fig. 5 (*4rd col*)).

In addition to the qualitative results presented above, quantitative evaluations have been performed using the intersection over union (IoU) as an evaluation metric of the obtained results; IoU is generally used to evaluate instance segmentation approaches (e.g., [25] and [5]). Table 3 presents the mean IoU values for the three approaches evaluated in this section computed using the three testing images presented in Fig. 5. Two IoU values are presented; the first one *(middle column)* corresponds to the IoU computed by considering the ground truth and segmentation result as a whole; in other words, the computed binary mask is evaluated over the ground truth, as a kind of global assessment. This evaluation does not take into account the accuracy of computed instances; hence, a second evaluation is performed element by element (note an element can be a corn kernel or an impurity). This second evaluation is presented in Table 3 *(right column)*. As can be appreciated, in both cases results from Mask R-CNN reach the highest performance.

Experiment	Mean IoU binary mask	Mean IoU per instances
Watershed	0.751	0.553
U-Net+Watershed	0.767	0.705
Mask R-CNN	**0.903**	**0.890**

Table 3. Results of segmentation stage.

Network	Good corn	Defective corn	Avg.Acc	# of Net. Param.
Mask R-CNN	0.962	0.644	0.803	63738 K
VGG16	0.950	0.917	0.933	134268 k
ResNet50	0.906	0.917	0.911	23591 K
CK-CNN	**0.956**	**0.933**	**0.945**	**3306 K**

Table 4. Results of classification stage for 2-classes.

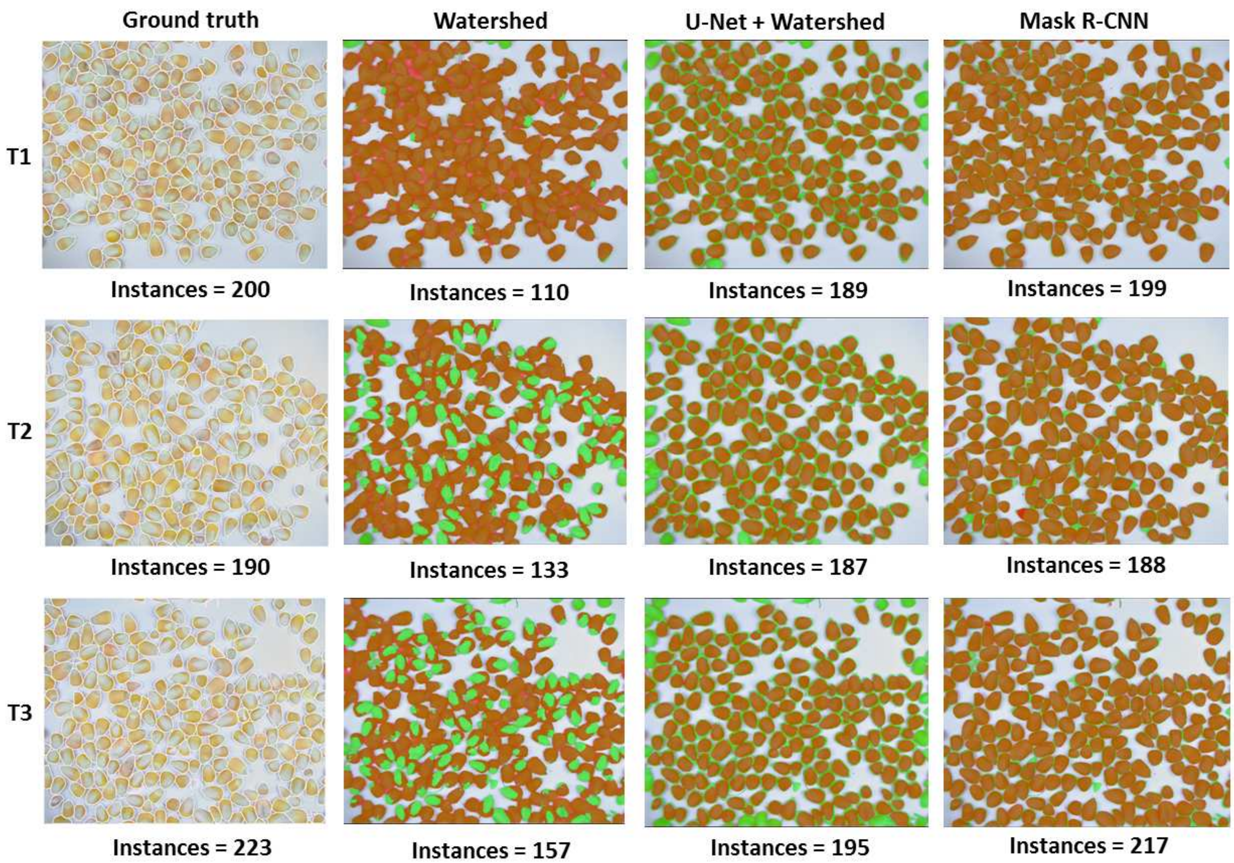

Figure 5. The total number of instances shown in each image represents the sum of good corn kernels, defective corn kernels and impurities. (*1st col*) Annotated image mask used as ground truth. (*2nd col*) Segmentation result obtained from watershed algorithm. (*3rd col*) Segmentation result from U-Net and watershed. (*4th col*) Segmentation result from Mask R-CNN.

Network	Good corn	Def. corn	Impurity	Avg. Acc	# of Net. Param.
Mask R-CNN	0.960	0.695	0.286	0.647	63738 K
VGG16	0.974	0.876	0.819	0.890	134272 k
ResNet50	0.986	0.860	0.931	0.925	23593 K
CK-CNN	**0.979**	**0.900**	**0.973**	**0.956**	**3306 K**

Table 5. Results of classification stage for 3-classes.

4.2. Classification

This section presents results obtained with the approach proposed for classifying the instances obtained with the previous segmentation process. As mentioned in Section 3.3, the 2-class (good corn and defective corn) and 3-class (good corn, defective corn and impurity) classification problems have been considered in the current work. Additionally, the fine-tuning of three architectures (i.e., VGG16, ResNet50 and Mask R-CNN) have been considered for quantitative comparisons.

For the 3-class classification, the proposed approach has been trained using 3600 images (1440 of good corns, 1440 of defective corns and 720 of impurities) and validated with 900 images (360 of good corns, 360 defective corns and 180 of impurities). Resulting in a total of 4500 images. This original set has been enlarged through a data augmentation process, which expands the given set by five times its size (i.e., vertical and horizontal flips, random rotations and width/height shift operations). Additionally, a set of 2100 images (700 images per each category) was kept aside for the testing stage. The above-mentioned dataset has been used for tackling the 2-class classification problem; in this case, 2310 images have been used for the training stage (1155 of good corns and 1155 of defective corns) and 990 images for the validation. The same data augmentation process has been considered, resulting in a total of 16500 labeled images. In the 2-class classification problem, 360 images have been considered for the testing stage. The training processes for the 3-class and 2-class problems have been performed through 300 epochs and 40 steps per epoch, both for our proposed model and the other architectures. The results obtained with the proposed novel lightweight network

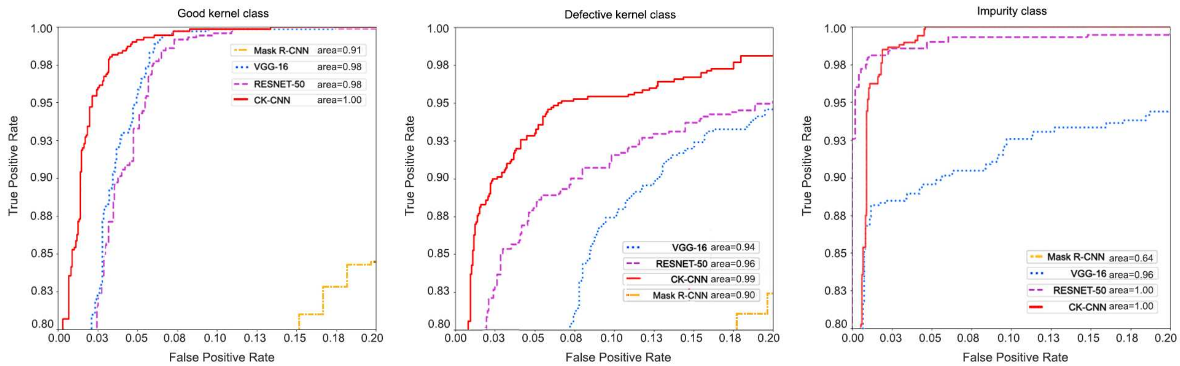

Figure 6. Comparisons of ROC curves for the 3-class-classification, CK-CNN vs other architectures (just an enlargement of the top left area of ROC curves is depicted).

CK-CNN are presented in Table 4 and Table 5, 2-class and 3-class respectively.

Regarding the comparisons, the first approach was the Mask R-CNN model, which has been used during the segmentation stage. The Mask R-CNN has been selected since it provides instance segmentation and classification at once. Unfortunately, although good results have been obtained in both cases, 2-class and 3-class problems, for the good corn category, it was not the case for the other categories. This low accuracy values in most of the classes are mainly due to the fact of the unbalance of classes present in the cluster set used for the training. For instance, looking at Table 2 we can appreciate the defective kernel and impurity classes have just a few instances in comparison with the good corn kernel category. This imbalance results in the poor discriminant capability of the network for these categories.

The other approaches used for comparisons are VGG16 and ResNet50 (see Section 2.2 for more details). In both cases, a pre-trained network has been considered and a fine-tuning process performed with the images from the current work. In more detail, in the case of VGG16, the first 14 layers have been maintained, while the last two layers have been retrained. In the case of ResNet50, all the network's parameters have been fine-tuned by using the dataset of the current work. On the contrary to the previous comparisons, in these cases, good results have been obtained in the two cases (2-class and 3-class classification problems), although the proposed approach reaches the best average performance in both of them. Just in the good corn category, for the 3-class classification, the ResNet50 architecture reaches a slightly better result (a 0.7% of improvement in accuracy). It should be highlighted that the number of parameters of ResNet50 is more than 7 times the number of parameters of CK-CNN; this difference is even bigger in the case of VGG16, in this case, the number of parameters is more than 40 times the number of parameters of CK-CNN. These results demonstrate that too deep learning models do not necessarily give us the best results in the particular case of kernel classification. The reduced number of parameters allows a fast convergence during the training process. To conclude these comparisonsFig. 6 present the ROC curves of the proposed approach (CK-CNN) together with the other three architectures (due to space limitations just the top left of ROC curves are depicted). It can be appreciated that the proposed CK-CNN architecture has not only the best quantitative results, with the lowest number of parameters (see Table 5) but also it has the best behavior.

5. Conclusions

This paper proposes a novel framework to classify elements from a given corn kernel sample set. The proposed approach is robust enough that no additional requirements are imposed on the way samples are distributed (i.e., no matter whether corn kernels are touching or separated). The proposed scheme consists of two phases; the first one is focused on the object segmentation, which is performed through the Mask R-CNN network. During the first phase instances on the given samples are extracted. During the second phase, each instance is classified using a lightweight network specially designed for this task. Experimental results are provided showing both the robustness of the proposed approach as well as the improvements on global classification performance with the proposed network architecture. As a second contribution of this work, a dataset has been generated and carefully annotated (single contour of corn kernels and class labels).

Acknowledgements

This work has been partially supported by the ESPOL project PRAIM (FIEC-09-2015); the Spanish Government under Project TIN2017-89723-P; and the “CERCA Programme / Generalitat de Catalunya”.

References

[1] Waleed Abdulla. Mask R-CNN for object detection and instance segmentation on keras and tensorflow. https://github.com/matterport/Mask_RCNN, 2017. 4

[2] Xiao Chen, Yi Xun, Wei Li, and Junxiong Zhang. Combining discriminant analysis and neural networks for corn variety identification. *Computers and electronics in agriculture*, 71:S48–S53, 2010. 3

[3] Pierre Dubosclard, Stanislas Larnier, Hubert Konik, Ariane Herbulot, and Michel Devy. Automated visual grading of grain kernels by machine vision. In *Twelfth International Conference on Quality Control by Artificial Vision 2015*, volume 9534, page 95340H. International Society for Optics and Photonics, 2015. 1, 2

[4] M Effendi, M Jannah, and U Effendi. Corn quality identification using image processing with k-nearest neighbor classifier based on color and texture features. In *IOP Conference Series: Earth and Environmental Science*, pages 012–066. IOP Publishing, 2019. 3

[5] P Ganesh, K Volle, TF Burks, and SS Mehta. Deep orange: Mask r-cnn based orange detection and segmentation. *IFAC-PapersOnLine*, 52(30):70–75, 2019. 6

[6] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017. 2, 3, 4, 6

[7] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015. 1, 3, 5

[8] Carlos Hernández, Ysamar Rodríguez, and Zulay Niño. Efecto del almacenamiento de granos de maiz (zea mays) sobre la calidad del aceite extraido. *Información tecnológica*, 20(4):21–30, 2009. 1

[9] Jeremiah W Johnson. Adapting mask-rcnn for automatic nucleus segmentation. *arXiv preprint arXiv:1805.00500*, 2018. 3

[10] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012. 1, 3

[11] Xiaoming Li, Baisheng Dai, Hongmin Sun, and Weina Li. Corn classification system based on computer vision. *Symmetry*, 11(4):591, 2019. 1, 2, 3

[12] Efthimia Mavridou, Eleni Vrochidou, George A Papakostas, Theodore Pachidis, and Vassilis G Kaburlasos. Machine vision systems in precision agriculture for crop farming. *Journal of Imaging*, 5(12):89, 2019. 1

[13] Chao Ni, Dongyi Wang, Robert Vinson, Maxwell Holmes, and Yang Tao. Automatic inspection machine for maize kernels based on deep convolutional neural networks. *Biosystems engineering*, 178:131–144, 2019. 1, 2, 3

[14] Giorgia Orlandi, Rosalba Calvini, Giorgia Foca, and Alessandro Ulrici. Automated quantification of defective maize kernels by means of multivariate image analysis. *Food Control*, 85:259–268, 2018. 1

[15] Diego Inácio Patrício and Rafael Rieder. Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review. *Computers and electronics in agriculture*, 153:69–81, 2018. 1

[16] Yibo Qin, Wei Wang, Wei Liu, and Ning Yuan. Extended-maxima transform watershed segmentation algorithm for touching corn kernels. *Advances in Mechanical Engineering*, 5:268046, 2013. 3

[17] Sergio Silva Ribeiro and JingTao Yao. Toward a three-way image classification model: A case study on corn grain images. In *2019 IEEE International Symposium on Multimedia (ISM)*, pages 177–1776. IEEE, 2019. 3

[18] Jos Roerdink and A. Meijster. The watershed transform: Definitions, algorithms and parallelization strategies. *FUNDINF: Fundamenta Informatica*, 41, 01 2000. 3

[19] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 3, 5

[20] Kate Sendin, Paul J Williams, and Marena Manley. Near infrared hyperspectral imaging in quality and safety evaluation of cereals. *Critical reviews in food science and nutrition*, 58(4):575–590, 2018. 2

[21] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 1, 3, 5

[22] Carmen Wallays, Bart Missotten, Josse De Baerdemaeker, and Wouter Saeys. Hyperspectral waveband selection for on-line measurement of grain cleanness. *Biosystems Engineering*, 104(1):1–7, 2009. 2

[23] Chao Xia, Sai Yang, Min Huang, Qibing Zhu, Ya Guo, and Jianwei Qin. Maize seed classification using hyperspectral image coupled with multi-linear discriminant analysis. *Infrared Physics & Technology*, 103:103077, 2019. 2

[24] Pavel Yakubovskiy. Segmentation models. https://github.com/qubvel/segmentation_models, 2019. 5

[25] Yang Yu, Kailiang Zhang, Li Yang, and Dongxing Zhang. Fruit detection for strawberry harvesting robot in non-structural environment based on mask-rcnn. *Computers and Electronics in Agriculture*, 163:104846, 2019. 3, 6

[26] Hua Zhu, Jun Yue, Zhenbo Li, and Zhiwang Zhang. Multi-kernel dictionary learning for classifying maize varieties. *International Journal of Agricultural and Biological Engineering*, 11(3):183–189, 2018. 3