

Incremental Two-Network Approach to Develop a Purity Analyzer System for Canola Seeds

Kuldeep Singh

Department of Electronics and Communication
Malaviya National Institute of Technology
Jaipur, India
Email: kuldeep.ece@mnit.ac.in

Fernando Saccon and Dileepan Joseph*

Department of Electrical and Computer Engineering
University of Alberta, Edmonton, AB
Canada T6G 1H9
*Email: dil.joseph@ualberta.ca

Abstract—Given a suitable dataset, transfer learning using deep convolutional neural networks is an effective method to develop a system to detect and classify objects. Despite having models pretrained on large general-purpose datasets, the requirement to manually label an application-specific dataset remains a limiting factor in system development. We consider this wider problem in the context of the purity analysis of canola seeds, where end users wish to distinguish species of interest from contaminants in images taken with optical microscopes. We use a Detector network, trained only to detect seeds, to help label the dataset used to train an Analyzer network, capable of both seed detection and classification. We present results, over three experiments that involve 25 contaminant species, including Primary and Secondary Noxious Weed Seeds (as per the Canadian Weed Seeds Order), to validate our incremental approach. We also compare the proposed system to competing ones in a literature review.

Index Terms—Artificial intelligence; neural network; object recognition; content detection; canola seeds; image analysis.

I. INTRODUCTION

Automated analysis and sorting systems have been developed [1], [2], [3], [4], [5], [6], [7], [8] for mm-sized objects, such as seeds, beans, and kernels, in the agriculture and food sector. Instead of artificial neural networks (ANNs), some works favour non-ANN image processing [2], [3], [6] or signal processing with multi-spectral sensors [8]. A relative advantage of these approaches is the analysis efficiency. Other works use ANNs but not convolutional neural networks (CNNs) [4], [5], even though such deep-learning ANNs have launched a revolution in object classification [9]. Because this revolution was still underway at the time it was written, a review paper on seed inspection methods [7] does not even address CNNs. A relative advantage of the ANN approaches, whether CNN or not, is the analysis accuracy.

Heo *et al.* [1] do exploit CNNs for seed analysis and sorting. Their work, the most relevant at the time of this research, combines non-ANN image processing for object detection, i.e., image segmentation, with CNNs for object classification. However, they do not take advantage of developments in semantic segmentation [10], whereby CNNs are used for both object detection and classification [11], [12], [13]. Also, like other authors, Heo *et al.* do not employ magnification. In contrast, Jahnke *et al.* [6] emphasize the importance of the objective lens, which they coupled to a camera, in a

system they designed to analyze canola seed (*B. napus*) and contaminants. Jahnke *et al.* also relied on a complex imaging setup that involved a pick-and-place robot.

In addition to this paper's relevance to its application, namely seed analysis and sorting, it is also significant to the computer vision discipline of object detection and classification using CNNs, which falls under image understanding. We propose an incremental two-network approach to reduce and even eliminate manually-labelled bounding boxes (MLBBs). A proposed Detector network is trained, validated, and tested using some images. Subsequently, we use it to generate automatically-labelled bounding boxes (ALBBs) for additional images, where objects in each image belong to one class only. An Analyzer network is then trained on all of these images to perform object analysis on new ones.

From the multitude of CNN architectures that are available, we restrict the scope of our *Applicability and Analysis* experiments to the single-shot detector (SSD), faster region-convolutional neural network (FRCNN), and region-based fully convolutional network (RFCN). The former two architectures are featured in a review paper by Forson [14]. As for the RFCN architecture, we chose it because it was faster during training and inference than the FRCNN one [11]. As it also outperforms the other two in terms of accuracy, we restrict our attention to just the RFCN architecture for our *Reduction of Effort* experiments.

II. APPLICABILITY AND ANALYSIS

We present our approach using three experiments: F0, K1, and K2. In this section, we introduce our dataset and focus on the F0 experiment. This experiment demonstrates the applicability of a selected CNN architecture, as well as four proposed metrics, for the fully-automatic detection and classification of all mm-sized objects in an optical microscopy image. Together with qualitative analysis, we propose a quantitative figure of merit (FoM) to establish a baseline for the *Reduction of Effort* experiments, K1 and K2, presented in the next section. The objects we use are canola seeds and contaminants.

In Table I, we divide the species of seeds in our dataset into four classes, as per the Canadian Weed Seeds Order [15] and a Canola Council of Canada FAQ [16]. Common names

are taken, firstly, from the US Department of Agriculture's PLANTS Database [17] and, secondly, from Wikipedia.

Figure 1 presents sample images of all species listed in Table I. The nine species considered in the F0 and K1 experiments are grouped into three classes, i.e., Classes 3, 6, and 7, whereas the K2 experiment has 18 additional species grouped into four classes, i.e., Classes 2, 3, 6, and 7. We compute what we call *class-based* results. This means that classifications are considered incorrect only if a species that belongs to one class is assigned a different class.

Table II indicates how one dataset of 315 images, collected incrementally, is used in three experiments. The F0 experiment uses the first 45 images, comprising 1,107 seeds. This experiment involves MLBBs only. The K1 and K2 experiments are discussed in the next section. (In addition to MLBBs, they involve ALBBs.) Though it may be done by a non-specialist, MLBB assignment requires substantial human effort for the large number of objects in a dataset suitable for the training, validation, and testing of a deep CNN for object detection and classification. In our dataset, we had a total of 2,893 mm-sized objects, i.e., canola seeds and contaminants.

We proposed a FoM to account not only for the classification errors but also for the detection errors. One can take for granted the ability of a human specialist to detect foreground objects perfectly from background ones. Even a non-specialist can detect seeds in an optical microscopy image. As we should not assume a computer can, we proposed four *metrics*, which we called Sensitivity, Accuracy, Specificity, and Precision, to help compute and explain the FoM:

$$\text{Sensitivity} = \frac{\text{True Detections}}{\text{Total Seeds}}, \quad (1)$$

$$\text{Accuracy} = \frac{\text{True Classifications}}{\text{Total Seeds} + \text{Extra Detections}}, \quad (2)$$

$$\text{Specificity} = \frac{\text{True Detections}}{\text{Total Seeds} + \text{Extra Detections}}, \quad (3)$$

$$\text{Precision} = \frac{\text{True Classifications}}{\text{Total Seeds}}. \quad (4)$$

Each metric ranges from 0 to 100%, where 100% is ideal.

The metrics of Equations 1 to 4 use several *measures*: True Detections, Total Seeds, True Classifications, and Extra Detections. Apart for the last one, they are basically self-explanatory. To count the Extra Detections, we first calculate the intersections-over-unions (IoUs), using the areas covered in pixels, of the nearest MLBBs and the ALBBs, i.e., the bounding boxes after object detection, as follows:

$$\text{IoU} = \frac{\text{MLBB} \cap \text{ALBB}}{\text{MLBB} \cup \text{ALBB}}. \quad (5)$$

The Extra Detections comprise Extra *True* Detections, where the IoU is greater than or equal to 50%, and Extra *False* Detections, where the IoU is less than 50%. Examples are given in Figure 2. As indicated, exactly one of the Extra True Detections is counted *instead* as the one True Detection.

While Sensitivity is the bottleneck in Equations 1 to 4, Accuracy is limited also by Specificity and Precision, which

offer additional insight into observed inaccuracies. This can be seen from the following mathematical relations:

$$\text{Accuracy} \leq \text{Sensitivity}, \quad (6)$$

$$\text{Specificity} \leq \text{Sensitivity}, \quad (7)$$

$$\text{Precision} \leq \text{Sensitivity}, \quad (8)$$

$$\text{Accuracy} \leq \text{Specificity}, \quad (9)$$

$$\text{Accuracy} \leq \text{Precision}. \quad (10)$$

To define the FoM, we use the product of Sensitivity and Accuracy or, what turns out to be equal, the same of Specificity and Precision. We call this FoM the Analysis Score:

$$\text{Analysis Score} = \text{Sensitivity} \times \text{Accuracy}, \quad (11)$$

$$= \text{Specificity} \times \text{Precision}. \quad (12)$$

As it proves to be a limiting factor, we propose the Specificity could be increased, in future work, by post-processing CNN output to eliminate Extra True Detections. This would raise the Accuracy, thereby raising the Analysis Score.

Table III lists the Analysis Scores of Class 7 for multiple scenarios and Monte Carlo trials. Class 7 has the species of interest, i.e., *B. napus* and *B. rapa*, for our Purity Analyzer System, which initially used a one-network approach to perform object detection *and* classification. Each trial represents a random partitioning of the available images into training, validation, and test sets. For every five images, three images (60%) are assigned to the training set, one image (20%) to the validation set, and one image (20%) to the test set.

We focus on the RFCN architecture in subsequent experiments, one of the three deep CNN architectures we evaluated for the Purity Analyzer System, because it outperformed both the FRCNN and the SSD architectures on median Analysis Score. This is visible in Table III. As an example, the 86.36% value, in the table, is calculated as follows:

$$\text{Analysis Score} = 100.00\% \times 86.36\%, \quad (13)$$

$$= 97.73\% \times 88.37\%, \quad (14)$$

$$(15)$$

where

$$\text{Sensitivity} = \frac{43}{43} = 100.00\%, \quad (16)$$

$$\text{Accuracy} = \frac{38}{43 + 1} = 86.36\%, \quad (17)$$

$$\text{Specificity} = \frac{43}{43 + 1} = 97.73\%, \quad (18)$$

$$\text{Precision} = \frac{38}{43} = 88.37\%. \quad (19)$$

These numbers correspond to the median RFCN result for Class 7 seeds in the F0 experiment.

Although we emphasize the Analysis Score in this paper, we examined metrics and measures besides those used by our proposed FoM. For example, with the RFCN network in the F0 experiment, a standard confusion matrix shows that Class 7 seeds were misclassified primarily as Class 3 seeds.

TABLE I
 CLASSES AND SPECIES OF SEEDS IN THE DATASET. THE CANADIAN WEED SEEDS ORDER MEANINGFULLY GROUPS SPECIES INTO CLASSES. CLASSES 2 AND 3 ARE PRIMARY AND SECONDARY NOXIOUS WEED SEEDS, RESPECTIVELY. WHEREAS CLASS 6 SPECIES ARE ALSO WEEDS (CONTAMINANTS), CLASS 7 GROUPS SPECIES (CANOLA) OF INTEREST TO END USERS (PURITY ANALYSTS) OF THE PURITY ANALYZER SYSTEM.

Class	Species	Common Name	Experiment
2	<i>Cirsium arvense</i>	Canada thistle	K2 only
2	<i>Datura stramonium</i>	Jimsonweed	K2 only
2	<i>Raphanus raphanistrum</i>	Wild radish	K2 only
3	<i>Daucus carota</i>	Wild carrot	K2 only
3	<i>Galium spurium</i>	False cleavers	F0, K1, K2
3	<i>Silene noctiflora</i>	Night-flowering catchfly	K2 only
3	<i>Sinapis arvensis</i>	Wild mustard	F0, K1, K2
3	<i>Thlaspi arvense</i>	Stinkweed	K2 only
3	<i>Vaccaria hispanica</i>	Cow cockle	F0, K1, K2
6	<i>Amaranthus retroflexus</i>	Redroot amaranth	K2 only
6	<i>Axyris amaranthoides</i>	Russian pigweed	F0, K1, K2
6	<i>Bassia scoparia</i>	Burningbush	F0, K1, K2
6	<i>Brassica juncea</i>	Brown mustard	K2 only
6	<i>Brassica oleracea</i>	Wild cabbage	K2 only
6	<i>Camelina sativa</i>	False flax	K2 only
6	<i>Chenopodium album</i>	Lambsquarters	F0, K1, K2
6	<i>Crepis tectorum</i>	Narrowleaf hawkbeard	K2 only
6	<i>Dracocephalum parviflorum</i>	American dragonhead	K2 only
6	<i>Fallopia convolvulus</i>	Black bindweed	K2 only
6	<i>Neslia paniculata</i>	Ballmustard	K2 only
6	<i>Persicaria lapathifolia</i>	Curlytop knotweed	K2 only
6	<i>Plantago major</i>	Broadleaf plantain	K2 only
6	<i>Raphanus sativus</i>	Wild radish	K2 only
6	<i>Sinapis alba</i>	White mustard	K2 only
6	<i>Solanum nigrum</i>	Black nightshade	F0, K1, K2
7	<i>Brassica napus</i>	Argentine canola	F0, K1, K2
7	<i>Brassica rapa</i>	Polish canola	F0, K1, K2

TABLE II
 NUMBERS OF IMAGES AND SEEDS IN THE DATASET. THE F0 EXPERIMENT IS DISCUSSED UNDER *Applicability and Analysis*. THE K1 AND K2 EXPERIMENTS ARE DISCUSSED UNDER *Reduction of Effort*, WHICH HAS *Additional Images* AND *Additional Species* SECTIONS.

Class	Species	Images			Seeds		
		F0	K1	K2	F0	K1	K2
2	<i>C. arvense</i>			+10		+60	
2	<i>D. stramonium</i>			+10		+70	
2	<i>R. raphanistrum</i>			+10		+41	
3	<i>D. carota</i>			+10		+28	
3	<i>G. spurium</i>	5	+10		106	+64	
3	<i>S. noctiflora</i>			+10		+66	
3	<i>S. arvensis</i>	5	+10		133	+66	
3	<i>T. arvense</i>			+10		+87	
3	<i>V. hispanica</i>	4	+10		79	+82	
6	<i>A. retroflexus</i>			+10		+119	
6	<i>A. amaranthoides</i>	4	+10		98	+63	
6	<i>B. scoparia</i>	4	+10		94	+60	
6	<i>B. juncea</i>			+10		+63	
6	<i>B. oleracea</i>			+10		+67	
6	<i>C. sativa</i>			+10		+112	
6	<i>C. album</i>	7	+10		244	+102	
6	<i>C. tectorum</i>			+10		+56	
6	<i>D. parviflorum</i>			+10		+45	
6	<i>F. convolvulus</i>			+10		+49	
6	<i>N. paniculata</i>			+10		+50	
6	<i>P. lapathifolia</i>			+10		+62	
6	<i>P. major</i>			+10		+68	
6	<i>R. sativus</i>			+10		+41	
6	<i>S. alba</i>			+10		+52	
6	<i>S. nigrum</i>	4	+10		93	+72	
7	<i>B. napus</i>	5	+10		94	+69	
7	<i>B. rapa</i>	7	+10		166	+72	
Subtotal:		45	+90	+180	1,107	+650	+1,136
Total:		45	135	315	1,107	1,757	2,893



Fig. 1. Sample images of all species in the dataset. These images correspond, from left to right and then top to bottom, to the 27 species listed in Table I, from top to bottom. Two different optical systems were used, which accounts for some variability in the backgrounds and foregrounds of images.

TABLE III
SELECTED ANALYSIS SCORES (100% RELATIVE EFFORT). UNDER *Applicability and Analysis*, THREE ARCHITECTURES ARE EVALUATED. AS THE REGION-BASED FULLY CONVOLUTIONAL NETWORK (RFCN) OUTPERFORMED THE SINGLE-SHOT DETECTOR (SSD) AND FASTER REGION-CONVOLUTIONAL NEURAL NETWORK (FRCNN), IT IS THE ONLY ARCHITECTURE USED UNDER *Reduction of Effort*.

Exp.	ConvNet	Species	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5
F0	SSD	<i>B. napus</i>	66.48	0.00	0.00	<u>26.32</u>	27.37
F0	SSD	<i>B. rapa</i>	50.17	<u>39.39</u>	84.03	27.04	34.02
F0	SSD	Class 7	57.11	17.02	40.88	26.72	<u>31.40</u>
F0	FRCNN	<i>B. napus</i>	89.47	52.00	30.00	84.76	<u>66.67</u>
F0	FRCNN	<i>B. rapa</i>	78.57	85.33	85.19	100.00	78.57
F0	FRCNN	Class 7	85.11	71.49	<u>82.98</u>	95.51	73.47
F0	RFCN	<i>B. napus</i>	94.74	51.59	37.06	<u>85.27</u>	89.47
F0	RFCN	<i>B. rapa</i>	80.00	100.00	<u>92.00</u>	100.00	85.19
F0	RFCN	Class 7	86.36	81.85	72.09	93.38	86.96
K1	RFCN	<i>B. napus</i>	64.52	71.34	<u>69.21</u>	75.00	57.20
K1	RFCN	<i>B. rapa</i>	65.22	100.00	<u>93.02</u>	97.30	87.15
K1	RFCN	Class 7	72.43	89.84	<u>86.02</u>	87.67	78.55
K2	RFCN	<i>B. napus</i>	42.35	<u>67.04</u>	53.63	<u>67.04</u>	79.93
K2	RFCN	<i>B. rapa</i>	97.67	<u>84.26</u>	82.05	69.92	88.59
K2	RFCN	Class 7	80.39	83.79	82.58	<u>82.85</u>	93.61

For each scenario in Table III, the median Analysis Score across trials is underlined. In addition to the Class 7 results, Analysis Scores are shown for *B. napus* and *B. rapa* when they are treated as individual species. We determined that it made most sense to focus this paper on class-based results. Meaningful aggregation of species, considering the numbers of seeds, results in more robust conclusions.

III. REDUCTION OF EFFORT

Having demonstrated the applicability of our Analyzer network, and introduced our method of analysis, we now turn our attention to the incremental development of a Purity Analyzer System using an incremental dataset. This involves an Analyzer network, which is incrementally retrained, and a Detector network. Together, the software and hardware comprise the Purity Analyzer System of Figure 3.

A. Additional Images

We consider two incremental steps. The first is the addition of new images, to the training, validation, and test sets, without changing seed species. This K1 experiment builds on Trial 4 of the F0 experiment detailed previously. We chose Trial 4 because, at the time we ran the K1 experiment, we were

focused on the median *B. napus* result. Only later, after completing the *Applicability and Analysis* work, did we switch to a class-based approach.

Using the initial F0 dataset, we trained a Detector network to automatically label all seeds in an image with bounding boxes. This deep CNN detects the seeds without assigning their class. It implements region-based binary classification using an identical architecture to the Analyzer network except at the output layer. The Detector network has fewer nodes and weights than the Analyzer network and yet must cope with the diverse appearance of seeds.

At each incremental step, we have n images per species. Suppose we trouble ourselves to assign an MLBB to each seed in only m of these images, where $0 \leq m \leq n$. If we compute ALBBs for the remaining images, the relative effort (RE) is m/n . Because MLBBs are precious, we assign the m images, first, to the training set and, second, to the validation set. Assigning them to the test set is least important because, unlike the other two sets, this set does not influence any weight of the Detector or Analyzer network.

For each Monte Carlo trial of the K1 experiment, the n images are in a specific random order. Only the first m images get MLBBs. Images are assigned to the training, validation,

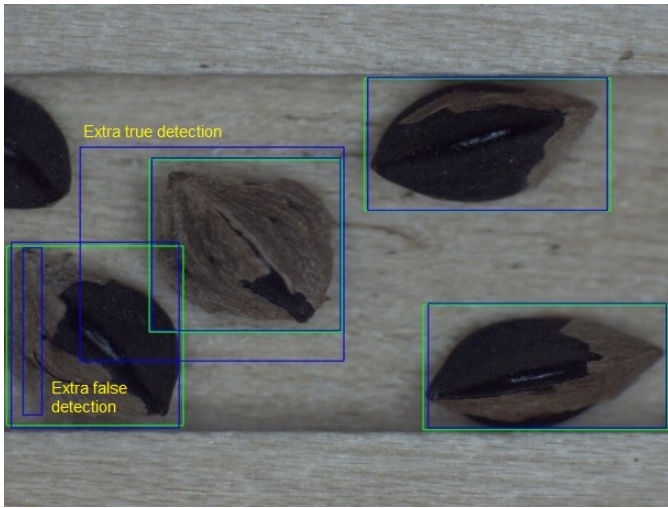


Fig. 2. Examples of True and Extra Detections. True detections are (blue) automatically-labelled bounding boxes (ALBBs) that overlap sufficiently with (green) manually-labelled bounding boxes (MLBBs). When a seed has multiple true detections, *all but one* are called Extra True Detections. The remaining *one* box is called the True Detection. An Extra False Detection is an ALBB that does not overlap sufficiently with any MLBB.

and test partitions in the same order. When m equals n , the incremental K1 dataset is partitioned and labelled in the same manner as was the initial F0 dataset during the *Applicability and Analysis* work.

Figure 4 illustrates, after the K1 experiment, that the *overall* Analysis Score value, computed over all classes of the F0&K1 test set, is consistently 80% or higher for all RE values. As with the F0 experiment, we reserved 20% of the K1 images for validation and 20% for test. Because n equals ten, there are two additional validation and two additional test images. The RE is varied by incrementing m from zero in steps of one, initially, and from six in steps of two, finally.

The K1 experiment involves additional images of nine species for which the Analyzer network was *already* trained. Class labels were easy to assign because each image comprised seeds of one class. To further validate our *two-network* approach to develop a Purity Analyzer System, we verify that it is beneficial to *retrain* the Analyzer network, using ALBBs assigned by the Detector network. For this purpose, Figure 4 includes a reference *one-network* result, computed using the original Analyzer network, on the same test set.

Considering the reference result, as shown in Figure 4, the K1 experiment demonstrates that our incremental two-network approach outperforms, especially at 0% RE, a non-incremental one-network approach. For the reference result, because there is no Analyzer retraining, or Detector training, only F0&K1 test-set data is required. At 100% RE, test-set results of the two-network approach are identical to those of a one-network approach retrained with the additional images.

B. Additional Species

We added to the dataset new images that comprise new species. This K2 experiment builds on Trial 3 of the K1 exper-

iment and Trial 4 of the F0 experiment. For each increment, we chose the trial that yielded the median *B. napus* result. Training, validation, and test proceed as before.

Figure 5 presents the test-set results of the K2 experiment. Although we manually labelled all seeds of the initial F0 dataset, and all seeds of the incremental K1 dataset, we need not manually label seeds of the incremental K2 dataset because we can create a Detector network from the combined F0&K1 dataset and use it to automatically label all of the incremental K2 dataset, additional test-set images included.

The incremental K2 dataset has species, including *Primary Noxious Weed Seeds* (Class 2), not present in either the initial F0 dataset or the incremental K1 dataset, i.e., the combined F0&K1 dataset. The initial F0, the combined F0&K1, and the combined F0&K1&K2 datasets all have *Secondary Noxious Weed Seeds* (Class 3), in addition to canola seeds (Class 7), whose purity matters to end users, and other contaminants (Class 6). Noxious Weed Seed and contaminant species are as listed in the Canadian Weed Seeds Order [15].

The Analysis Score is computed using a test set only. For the K1 and K2 experiments, we incorporated ALBBs into test sets to enable the automatic testing of incremental improvements to the Analyzer network. This idea worked consistently well for the *Additional Images* scenario (K1 experiment), summarized in Figure 4, and inconsistently well for the *Additional Species* scenario (K2 experiment), summarized in Figure 5.

For the *Additional Species* scenario, Figure 6 provides an Analysis Score using MLBBs in the test set, irrespective of the RE. As it demonstrates that the ground-truth Analysis Score, when test-set data uses MLBBs, is consistently higher than was otherwise thought (Figure 5), our hypothesis that automatic testing is less useful for this scenario is proven correct.

In the K2 experiment, no new Class 7 (canola class) seeds were added. The reason the Analysis Score of Class 7 differs, in Figures 5 and 6 compared to Figure 4, is that even though training, validation, and test sets use the same bounding boxes for Class 7 seeds, all MLBBs, detection performance of the Analyzer network on Class 7 is affected by retraining performed with the additional species. Moreover, the proportion of MLBBs for the additional species varies with RE.

Finally, we note that at 0% RE, i.e., with no MLBBs for the *Additional Species* images, our incremental two-network approach is able to detect and classify novel species. In terms of the end application, it is significant that the new species include Primary and Secondary Noxious Weed Seeds.

IV. CONCLUSIONS

We proposed a method to train, validate, and test a CNN for detection and classification of mm-sized objects using a dataset that is grown over time. Initially, there is an F0 dataset (45 images). Later, we add a K1 dataset (+90 images), yielding a combined F0&K1 dataset. Finally, we add a K2 dataset (+180 images), yielding a combined F0&K1&K2 dataset. The method may be generalized to other object datasets.

Although we manually labelled *all* the initial F0 dataset, we did not have to manually label *all* the incremental K1 dataset



Fig. 3. Prototype of a Purity Analyzer System. Software (illustrated left) includes a CNN, called the Analyzer network, for object detection and classification. Another CNN, called the Detector network (not shown), facilitates training, validation, and testing. Hardware (illustrated right) comprises a seed container (not shown), a custom dispenser, a mini conveyor, an optical microscope, a ring illuminator, a digital camera, and a desktop computer (not shown).

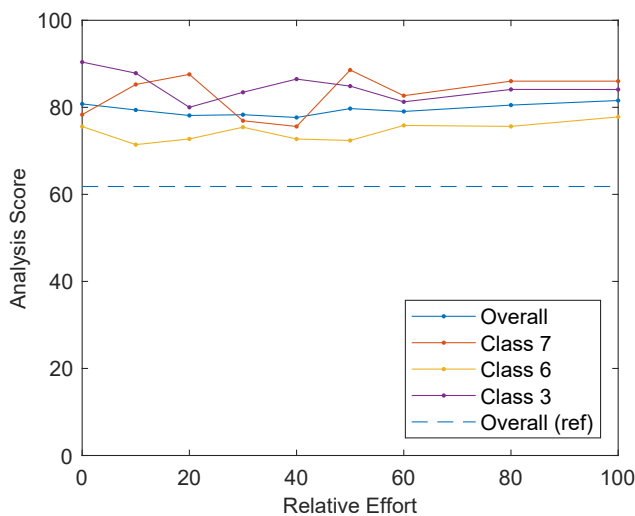


Fig. 4. Analysis Score vs. RE on the F0&K1 test set. At 0% relative effort (RE), all the additional K1 images use ALBBs. At 30% RE, half the additional K1 training images use MLBBs whereas, at 60% RE, all of them use MLBBs. At 80% RE, all the additional K1 training and validation images use MLBBs whereas, at 100% RE, all K1 images use MLBBs. The overall class-based Analysis Score is representative of each class-based Analysis Score.

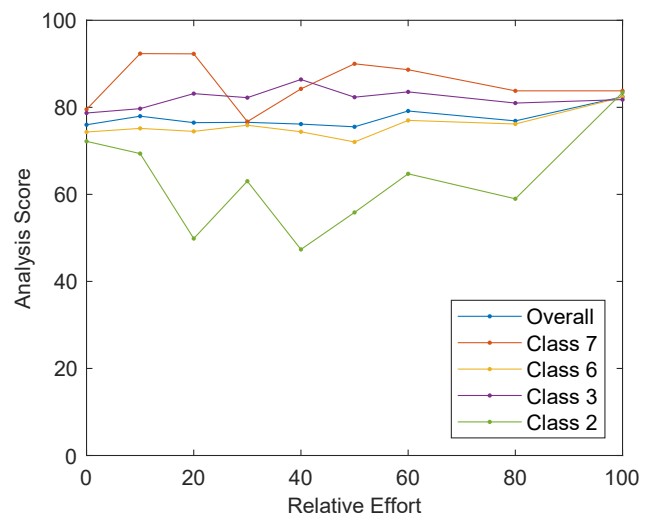


Fig. 5. Analysis Score vs. RE on the F0&K1&K2 test set. By varying which K2 images use ALBBs and which ones use MLBBs, the RE is varied. Perhaps because the additional K2 images comprise additional species, using ALBBs with the additional K2 test images proves less useful here – it lowers the Analysis Score of Class 2 especially, even though there is still a benefit to using ALBBs with additional Class 2 training and validation images.

because we created a Detector network, using *all* of the F0 and *none* (or *some*) of the K1 data, and leveraged it to train, validate, and test an Analyzer network, on *all* of the combined F0&K1 data (135 images of 1,757 seeds). A similar approach is taken when the K2 data is incrementally added.

The main significance of our two-network approach is a *Reduction of Effort*. With a Detector network, we automatically labelled the bounding boxes of *all* mm-sized objects in images taken with optical microscopes. These images are used to train, validate, and test an improved Analyzer network to not only detect but also to classify similar objects. Representing two different ways to reduce effort, our research included *Additional Images* and *Additional Species* experiments.

We investigated the impact of RE, defined as the ratio of

automatically-labelled to total bounding boxes, i.e., ALBBs to MLBBs plus ALBBs. Moreover, we introduced a FoM, i.e., the Analysis Score, to assess the impact of both classification *and* detection errors on overall system performance.

We conducted Monte Carlo analyses composed of five trials each, in which the partitioning of our dataset into training, validation, and test sets, and the ordering therein, was randomly varied. For each incremental experiment, we used the partition from a prior experiment that yielded a median Analysis Score. Consequently, the Monte Carlo analyses enhance the statistical robustness and, therefore, validity of our conclusions.

Our experiments featured 25 contaminant species, including Primary and Secondary Noxious Weed Seeds, as determined by the Canadian Weed Seeds Order. This also demonstrates

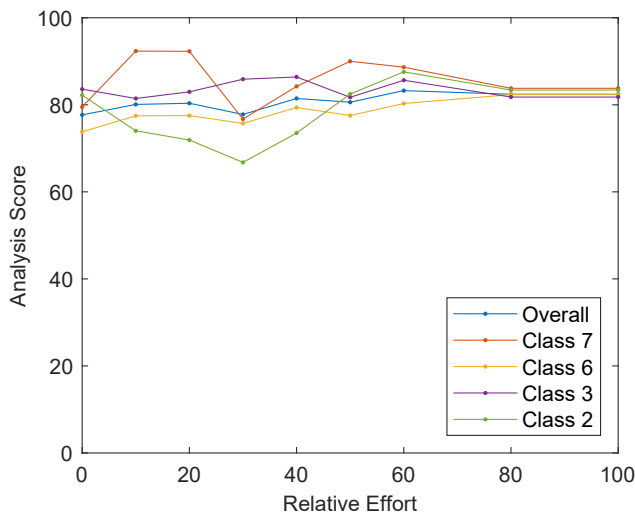


Fig. 6. Analysis Score vs. RE on the F0&K1&K2 test set. Whereas the experiment is otherwise identical to that of Figure 5, MLBBs are used for all seeds in the test-set images. The higher and more stable Analysis Score results, in this figure compared to the previous, confirms a benefit of the proposed incremental two-network approach even for additional species. Class 2 results correspond to previously-unseen Primary Noxious Weed Seeds.

significance, especially to potential end users from Canada, like the Purity Analysts with whom we engaged.

ACKNOWLEDGEMENTS

In alphabetical order, the authors thank Sarah Foster, Robin Harrish, Jill Himschoot, Kim Kenward, James Lam, and Carey Matthiessen of 20/20 Seed Labs. They also especially thank Kevin Zaychuk, formerly of 20/20 Seed Labs, and Kasturi Narayanan, of the National Research Council (NRC).

Moreover, the authors thank Canada’s Natural Sciences and Engineering Research Council (NSERC) for Partnership Grants (EGP 493529-16 and EGP2 508766-17) that supported this work. They also thank 20/20 Seed Labs, who offered in-kind contributions and funded this work with NSERC.

REFERENCES

- [1] Y. J. Heo, S. J. Kim, D. Kim, K. Lee, and W. K. Chung, “Super-High-Purity Seed Sorter Using Low-Latency Image-Recognition Based on Deep Learning,” *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 3035–3042, 2018.
- [2] L. Quan, T. Zhang, L. Sun, X. Chen, and Z. Xu, “Design and Testing of an On-Line Omnidirectional Inspection and Sorting System for Soybean Seeds,” *Applied Engineering in Agriculture*, pp. 1003–1016, 2018.
- [3] K. Kanjanawanishkul, P. Chupawa, and T. Nuantoon, “Design and assessment of an automated sweet pepper seed sorting machine,” *Engineering in Agriculture, Environment and Food*, vol. 11, no. 4, pp. 196–201, 2018.
- [4] K.-Y. Huang and J.-F. Cheng, “A Novel Auto-Sorting System for Chinese Cabbage Seeds,” *Sensors*, vol. 17, no. 4, p. 886, 2017.
- [5] F. Kurtulmus, I. Alibas, and I. Kavdir, “Classification of pepper seeds using machine vision based on neural network,” *International Journal of Agricultural and Biological Engineering*, vol. 9, no. 1, pp. 51–62, 2016.
- [6] S. Jahnke, J. Roussel, T. Hombach, J. Kochs, A. Fischbach, G. Huber, and H. Schar, “phenoSeeder – A Robot System for Automated Handling and Phenotyping of Individual Seeds,” *Plant Physiology*, vol. 172, no. 3, pp. 1358–1370, 2016.

- [7] Z. Gong, F. Cheng, Z. Liu, X. Yang, B. Zhai, and Z. You, “Recent Developments of Seeds Quality Inspection and Grading Based on Machine Vision,” in *ASABE Annual International Meeting*, pp. 1–16, American Society of Agricultural and Biological Engineers, 2015.
- [8] R. P. Haff, T. C. Pearson, and E. Maghirang, “A multispectral sorting device for isolating single wheat kernels with high protein content,” *Journal of Food Measurement and Characterization*, vol. 7, no. 4, pp. 149–157, 2013.
- [9] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [10] J. Long, E. Shelhamer, and T. Darrell, “Fully Convolutional Networks for Semantic Segmentation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431–3440, 2015.
- [11] J. Dai, Y. Li, K. He, and J. Sun, “R-FCN: Object Detection via Region-based Fully Convolutional Networks,” in *Advances in Neural Information Processing Systems*, pp. 379–387, 2016.
- [12] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, “SSD: Single Shot Multibox Detector,” in *European Conference on Computer Vision*, pp. 21–37, Springer, 2016.
- [13] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real-time Object Detection with Region Proposal Networks,” in *Advances in Neural Information Processing Systems*, pp. 91–99, 2015.
- [14] E. Forson, “Understanding SSD MultiBox – Real-Time Object Detection In Deep Learning,” *Towards Data Science*, 2017.
- [15] M. o. J. Canada, “Weed Seeds Order.” <https://laws-lois.justice.gc.ca/eng/regulations/SOR-2016-93/page-2.html>, 2016. SOR/2016-93.
- [16] C. C. o. Canada, “Canola FAQs.” <https://www.canolacouncil.org/oil-and-meal/what-is-canola/canola-faq's/>, 2020.
- [17] D. o. A. United States, “PLANTS Database.” <https://plants.usda.gov/java/>, 2019. Natural Resources Conservation Service.