Super-High-Purity Seed Sorter Using Low-Latency Image-Recognition Based on Deep Learning

Young Jin Heo, Se Jin Kim, Dayeon Kim, Keondo Lee, and Wan Kyun Chung, *Fellow, IEEE*

***Abstract*—Most commercial optical sorting systems are designed to achieve high throughput, so they use a naive low-latency image processing for object identification. These naive low-latency algorithms have difficulty in accurately identifying objects with various shapes, textures, sizes, and colors, so the purity of sorted objects is degraded. Current deep learning technology enables robust image detection and classification, but its inference latency requires several milliseconds; thus, deep learning cannot be directly applied to such real-time high throughput applications. We therefore developed a super-high purity seed sorting system that uses a low-latency image-recognition based on a deep neural network and removes the seeds of noxious weeds from mixed seed product at high throughput with accuracy. The proposed system partitions the detection task into localization and classification, and applies *batch inference only once* strategy; it achieved 500-fps throughput image-recognition including detection and tracking. Based on the classified and tracked results, air ejectors expel the unwanted seeds. This proposed system eliminates almost the whole weeds with small loss of desired seeds, and is superior to current commercial optical sorting systems.**

***Index Terms*—Agricultural automation, deep learning in robotics and automation, computer vision for automation.**

I. Introduction

AUTOMATED sorting systems physically separate specific objects from a mixture. This technology has a long history in many industries and research areas [1]–[4]. Especially, optical sorting systems that can separate millimeter-sized objects such as nuts, seeds, and beans have been widely used in many agricultural and food industries [5]–[8]. These sorting systems have mainly focused on removing contaminants from a product, while achieving high throughput. However, some industries require a cleaned product with high purity such as seeds, grains, and pharmaceutical tablets [8]–[10].

Optical sorting systems use optical cameras to detect small objects based on their color and morphology, then use an actuator such as air ejector to remove unwanted objects [8]. Most commercial optical sorting systems are designed to achieve high throughput (>1 ton/h), so they perform low-latency image processing to identify target objects. This approach has shortcomings in some applications. For example, seeds have various colors, shapes, textures and sizes depending on their kind. Also, variation in the pose of seeds can complicate their identification. Because of this diversity, naive (low-latency) image processing algorithms have limited accuracy when identifying the type of seed [11]; it is the main impediment to realization of sorting systems that achieve both high throughput and high purity.

Recently, image classification and detection using deep learning have achieved remarkably high accuracy [12]–[15]. A convolutional neural network (CNN) can extract features from images autonomously; this ability enables classification with high accuracy. Despite the good accuracy of deep learning, it cannot be directly applied to such a high-throughput optical sorting systems because most CNNs have high computation load, so they require large inference latency, which prevents achievement of high throughput.

In this study, we propose a real-time image-recognition method that has high accuracy and low latency. The method uses a CNN-based image classifier and a high-throughput inference strategy which is called *batch inference only once* strategy. The method partitions the object-detection task into localization and classification to decrease the inference latency while preserving classification accuracy. The CPUs perform object localization and tracking by using low-latency image processing and a Kalman filter motion tracker while the GPU performs CNN-based classification. The proposed system achieved 500-fps image-recognition throughput including object detection and tracking using a single camera and GPU. Based on the proposed high-throughput and high-accuracy image-recognition method, an optical seed-sorting system was developed and the system achieved both high throughput and high purity. The classification accuracy was 99.58% and achieved 99.994% purity of the desired seed, with only 1.5% loss; both results are superior to those of commercial high-throughput sorting systems. The object detection accuracies were evaluated quantitatively by comparing state-of-the-art detectors. Also, sorting accuracy

Manuscript received February 24, 2018; accepted June 4, 2018. Date of publication June 21, 2018; date of current version July 9, 2018. This letter was recommended for publication by Associate Editor F. Dayoub and Editor C. Stachniss upon evaluation of the reviewers' comments. This work was supported in part by the National Research Foundation of Korea grant funded by the South Korea government under Grant 2011-0030075 and in part by the Industrial Technology Innovation Program under Grant 10048358 funded by the Ministry Of Trade, Industry & Energy (MI, South Korea). This paper is part of IEEE Robotics and Automation Letters' Special Issue on Precision Agricultural Robotics and Autonomous Farming Technologies, edited by H. S. Ahn, I. Sa, and F. Dayoub. *(Corresponding author: Wan Kyun Chung.)*

The authors are with the Department of Mechanical Engineering, Pohang University of Science and Technology, Pohang-si 37673, South Korea (e-mail: heoyoungjin@postech.ac.kr; universe2030@postech.ac.kr; dayon95@postech.ac.kr; rjseh8405@postech.ac.kr; wkchung@postech.ac.kr).

This letter has supplemental downloadable multimedia material available at http://ieeexplore.ieee.org, provided by the authors. The Supplementary Materials contain a video showing a short presentation of the manuscript. The proposed image recognition system achieved 500-fps throughput image detection and tracking with accuracy based on deep learning technology. Images include validation mixed seed images with bounding boxes and classes predicted by the proposed image recognition pipeline. This material is 60 MB in size.

Digital Object Identifier 10.1109/LRA.2018.2849513

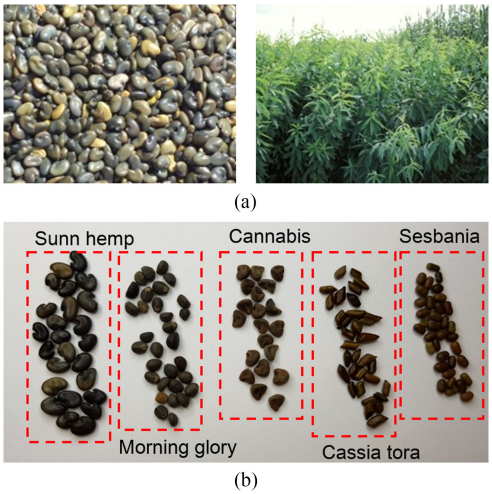

Fig. 1. (a) Sunn hemp seeds and its crops. (b) Sunn hemp and included noxious weeds.

of the proposed system was compared with commercial optical sorting systems in experiments that used the same target sorting objects.

The remainder of this letter is organized as follows: Section II presents background information and describes the configuration of the proposed system. Section III explains the seed-sorting strategy including fast image-recognition and sorting algorithm. Section IV presents experiments to evaluate the proposed high purity seed sorting system, and Section V presents discussion and conclusion.

II. Preliminaries and System Configuration

In this section, we specify the target object and provide an overview of the system. To achieve a sorting system that achieves super-high purity, we found a specific application which is required in the seed industry. Some background information will be provided to illustrate the challenges of a super-high-purity seed sorter.

A. Sunn Hemp Seeds and Noxious Weeds

Crotalaria juncea, usually called sunn hemp, is a tropical Asian plant and green manure used as a cover crop to revitalize poor land (Fig. 1(a)). Sunn hemp has many practical applications because it can produce a large amount of biomass in a short time, reduce soil erosion, and improve the fertility of soil [16]. However, many sunn hemp products include seeds of noxious weeds such as cannabis and morning glory, which must be removed to realize completely weed-free sunn hemp: in this study, a product with purity 99.99% is regarded as weed-free.

The number of seeds in 1 kg of sunn hemp is ∼24,500, and typically includes 300 seeds of noxious weeds, and wild beans. In this study, we use sunn hemp with seeds of four noxious weeds, and sort the seed into five categories: 1) sunn hemp, 2) morning glory, 3) cannabis, 4) Cassia tora, and 5) Sesbania (Fig. 1(b)). The objective of this study is to remove all weeds in mixed sample using developed sorting system.

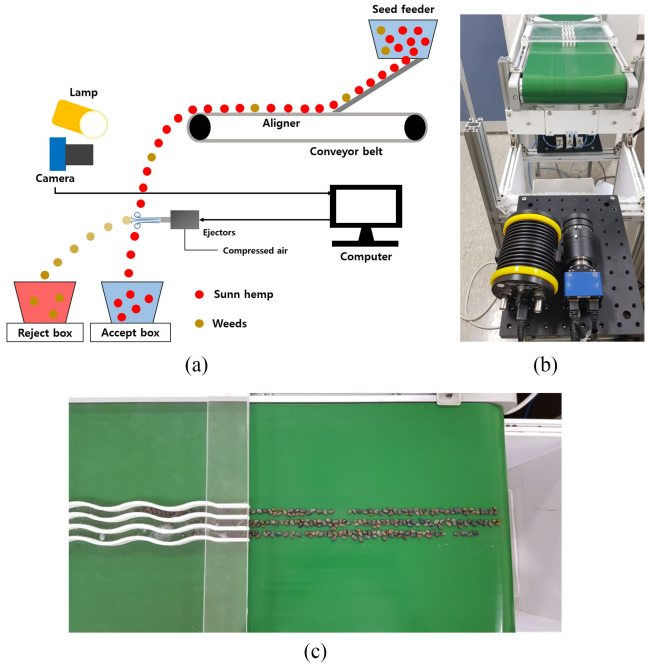

Fig. 2. (a) Overall structure of a belt-type seed sorting system (b) Front view of the testbed to evaluate the proposed image-recognition pipeline and sorting system. (c) Seed aligner that three channels on the conveyor belt generate three aligned seed streams.

B. Seed Sorter and Testbed

Many commercial optical sorters are composed of cameras, LED light, seed provider, and air ejectors (Fig. 2(a), (b)). The seed provider is composed of a feeder, a conveyor belt, and an aligner. The seed provider aligns inserted seeds and generates parallel single-seed streams from a pile of seeds to prevent seed overlap. After the seeds are aligned, they fall by gravity into an accept box under the conveyor belt. While the seeds fall, the image-acquisition system (i.e., several cameras; LED lights) obtains images of them in real-time and an image-processing system decides whether each seed should be accepted or rejected. Seed identification by an image-recognition algorithm is the most challenging problem because seeds rotate during free fall and have various colors, shapes, textures, and sizes (Fig. 3). When the image processing determines that a seed is unwanted, the air ejector blows it into the reject box.

We constructed a prototype to perform a sorting experiment and to evaluate the proposed image-recognition pipeline (Fig. 2(b)). The testbed consists of a camera, LED light, and three solenoid valves which can be regarded as a single *sorting module*. Usually, commercial sorting systems use several sorting modules in parallel to increase throughput. On the contrary, we concentrate on the image acquisition, recognition, and sorting parts to achieve weed-free sorting by greatly improving seed identification accuracy with high speed; i.e., we focus on high purity and high throughput in each sorting module, rather than high throughput of massive sorting systems. Our single sorting module is composed as follows.

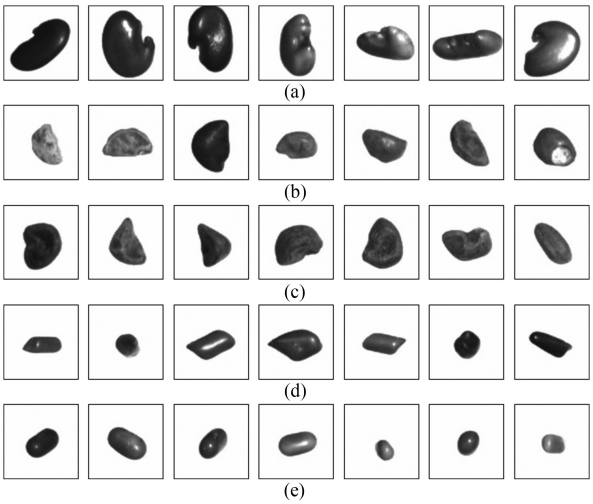

Fig. 3. Seed images acquired by CMOS camera while seeds fall by gravity. The five categories show various textures, sizes, morphologies, and colors. Even the same category has various shapes because seeds rotate while they fall. These factors significantly hinder the image-recognition that identifies the type of falling seeds. Each image has 100 × 100 pixel resolution; each 1 pixel is 0.08 mm. (a) Sunn hemp. (b) Moning glory. (c) Cannabis. (d) Cassia tora. (e) Sesbania.

1) Seed Provider: This apparatus is composed of a conveyor belt and an aligner. The conveyor belt makes seeds continuously fall to the floor. The seed aligner consists of three winding channels; while the seeds are transported on the conveyor belt, the aligner transforms a pile of seeds into three seed streams (Fig. 2(c)).

2) Image Acquisition: Images are acquired by a single CMOS camera (MATRIX-VISION mvBlueCOUGAR-XD) with an LED light source (VERITAS Constellation 120E); together these devices can capture high-definition images that clearly represent contrast, texture, and shape of seeds (Fig. 3); these factors can be used to identify the species of seeds. After considering the falling time, velocity of seeds in free fall, and resolution of single-seed image in the region of interest (ROI) image, we set frame rate to 500 fps and ROI size to 500 × 500 pixels. We set the distance between camera and seeds to ensure that a single seed occupied under 100 × 100 pixels in a image (Fig. 3). Here, the ROI is defined by field of view of the camera, and single-seed images will be the input of the classifier in the proposed pipeline. We can increase the throughput of a single camera by increasing the field of view, but this method has trade-offs: increasing the field of view image increases the number of seeds in an image, and thereby increases computational cost. A detailed analysis of this problem will be given in III-C.

3) Air Ejector: This device is composed of a digital controller, three solenoid valves, and three air nozzles. While seeds fall to floor, the air ejector shoots compressed air at the targeted floating seed and blows it to the reject box. The characteristics of the air nozzles determine the range and speed of the air jet; flow rate becomes high when cross-sectional area of the nozzle is small, and the width of the jet increases when the cross-sectional area of the nozzle is large. We optimized the cross-sectional areas of the three nozzles because each solenoid valve covers a specific region and shoots one seed at time.

TABLE I
CONFUSION MATRIX REPRESENTING CLASSIFICATION RESULT; "POSITIVE"=SUNN HEMP SEED, "NEGATIVE"=WEED SEED

		Predicted class	
		Positive	Negative
True class	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)
	Sorting	Accept	Reject

C. Purity and Loss

The effectiveness of the desired sorter can be quantitatively evaluated by purity and loss. The main purpose of the proposed sorting system is to realize high-purity and small loss to maximize profitability. The sorter separates heterogeneous seeds into two boxes: accept and reject (Fig. 2(a)). The purity is calculated from the accept box and the loss is calculated from the reject box (Table I) as

$$\text{purity} = \frac{P_{sh}\text{TP}}{P_{sh}\text{TP} + (1 - P_{sh})(1 - \text{TN})}, \tag{1}$$

$$\text{loss} = P_{sh}(1 - \text{TP}) + (1 - P_{sh})\text{TN}, \tag{2}$$

where P_{sh} is the initial sunn hemp ratio of mixed sample and is approximately $P_{sh} = 98.77\%$ (=100 × (24500−300)/24500) as mentioned in Section II-A. *True positive* (TP) is identification of sunn hemp as sunn hemp, *true negative* (TN) is identification of weed as weed, *false positive* (FP) is identification of weed as sunn hemp, and *false negative* (FN) is identification of sunn hemp as weed; all of these quantities are expressed as proportions, which sum to 1. TP and FP will go to the accept box and TN and FN will go to the reject box. Purity approaches 1 (weed-free) as TN approaches 1, so to achieve super-high purity, TN must be maximized. Loss converges to 0 as TP approaches 1, so to minimize loss, TP must be maximized.

D. Effectiveness of Commercial Sorters

We tested sorting capability of two commercial optical sorters: FMS2000-F (Satake Corporation) and BELTUZA (Satake Corporation). FMS2000-F is a chute type optical sorter with 2 full color RGB cameras and BELTUZA is a belt type optical sorter with 4 full color CCD and 2 NIR cameras.

To remove weed seeds completely, the initial mixed sample was sorted twice using the commercial sorters. During the first trial, the sorter operated at 20% of maximum sorting speed to improve sorting accuracy; the result was that 50% of the original mixed seeds were accepted and the accepted sample included many weed seeds. The second trial used the accepted mixed seeds from the first trial, and was performed at 5% of maximum sorting speed; about 45% of the initially-accepted seeds were accepted at the second time, and purity was $\sim 99.0\%$ (not weed-free). Two commercial sorters showed similar results. These results mean that the commercial sorters cannot achieve complete weed removal, and that they lose large amounts of sunn hemp seeds; resultant loss was about 27.5%. We consider that these weak results are caused by the limitation of naive low-latency image processing system that these commercial sorters use.

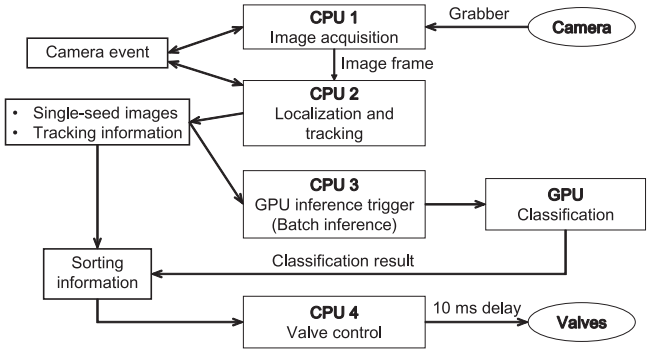

Fig. 4. Proposed image-recognition system for seed sorting. CPU 1 catches images from camera and sends them to CPU 2 every 2 ms. CPU 2 performs localization and tracking on the passed images and obtains single-seed images by exploiting the tracking information. CPU 3 triggers the GPU, which uses a trained network to predict the class of each single-seed image within 5–7 ms. CPU 4 uses the tracking and classification results to determine the timing of shooting air to eject weed seeds, and sends a control signal to the valves to do so.

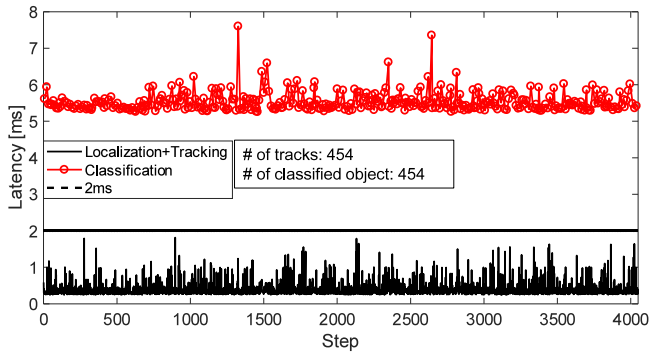

Fig. 5. Elapsed time of localization, tracking, and classification by the proposed pipeline. Required time for both localization and tracking never exceeds 2 ms, and classification needs 5–8 ms even if jittering is occurred. The number of tracks computed by the tracker is same as the number of objects classified by the classifier (454); i.e., all seeds were evaluated by the classifier and it yields 500-fps throughput.

III. Fast Image-Recognition System: Detection, Tracking, Classification, and Sorting

In this section, we introduce the developed image-processing system that enables detection, tracking, and sorting in real- time at 500 fps. The main contribution of this pipeline is that it enables use of a CNN-based classifier (ResNet-18) for a real-time high-throughput system in spite of its large inference latency (typically >5 ms). The main idea to achieve the contribution is that the separation of detection task into localization and classification tasks, and the classification is performed only once per object while the object passes through the ROI.

Four CPUs and a single GPU perform the assigned tasks in parallel (Fig. 4). CPU 1 acquires ROI images from the CMOS camera at 2 ms intervals (Section II-B2) and CPU 2 sequentially performs localization and tracking in every image frame (Section III-B). CPU 3 obtains cropped single-seed images, each with its identity (ID), from CPU 2 and triggers the GPU so that the classifier receives the single-seed images as an input and can infer the classes of each object (Section III-D). Here, the classifier requires 5–7 ms for GPU inference. However, *batch inference only once* strategy enables every object to be predicted by the classifier during the process with 500-fps (Section III-B). CPU 4 computes the timing of valve shooting by considering seed position estimated by using a Kalman filter tracker (Section III-E). Elapsed time during an experiment shows that localization and tracking performed its tasks within 2 ms and that classification was performed at single time while seeds passed through the ROI (Fig. 5).

A. Separating Detection as Localization and Classification

In the image-recognition field, *object-detection* is the process of identifying objects of a certain class; the object detector performs localization and classification concurrently. Localization finds bounding boxes of any number of semantic objects in an image; classification finds the class of the localized object. Modern object detectors use CNNs and have focused on developing end-to-end learning architectures that can directly extract bounding boxes with their classes [17]–[20]. State-of-the-art object detectors have become progressively faster, but to the authors' best knowledge, end-to-end detectors cannot yet achieve detection and tracking with 500-fps throughput for an input image that has 500 × 500 resolution.

To achieve 500-fps throughput for real-time sorting, we separate a detection task as localization and classification. The main requirement is to increase classification accuracy by applying a state-of-the-art classifier such as Inception and ResNet, which are based on a CNN [14], [15]. The environment of seed sorting system has a fixed background, and objects are in free fall; therefore, object localization can be easily accomplished by applying naive image processing that is computed quickly. Using the localization result (bounding box of each object), we can obtain single-seed images that will be the input of a CNN-based classifier (Fig. 3); i.e., to perform object-detection we apply a naive image processing for localization, then a CNN for classification. By separating localization and classification, they can be performed in parallel using CPUs and a GPU. Because this architecture is used, CNNs that are reliable but require large latency can be applied as a seed classifier to achieve real-time high-throughput sorting.

For the localization step, a single CPU thread performs a low-latency image processing. First, the pre-acquired background image is subtracted from an original image, and the subtracted image is binarized using a threshold. Then morphological closing (dilation followed by erosion) is applied to the binarized image to remove noise from the image. Then a blob detector finds the contours of each object; detected blobs that have larger area than a certain threshold are declared to be localized objects. The area thresholding removes outliers such as dust or shadow that should not be detected. The whole process takes <0.3 ms. The localized result is subject to multiple-object tracking.

B. Tracking and Batch Inference Only Once Strategy

Because seeds in free fall are represented as an image sequence, an object-tracking algorithm should be applied to find

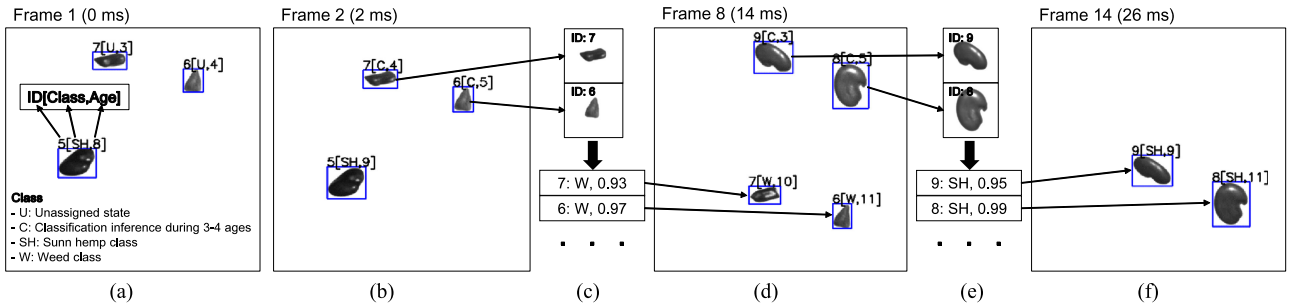

Fig. 6. *Batch inference only once* strategy for high-throughput deep learning inference. (a) Each object track consists of ID, class, and age. Here, class consists of unassigned state (U), during inferring state (C), sunn hemp (SH), and weed (W). (b) When the age of an object with unassigned state becomes 5, classification of tracks under age 5 is performed during frame ages 3–4. (c) After classification is complete, the predicted class and its probability are returned. (d) The classification results are assigned to its corresponding tracks. At the same time, tracks that newly become age 5 or less are subject to the next classification, as in (b). (e) and (f) repeat the same processes as (c) and (d) continuously.

corresponding seeds in consecutive frames. Moreover, the object tracker enables calculating of accurate ejection timing for physical sorting (Section III-E). We used a conventional Kalman filter motion tracker, which uses a constant-acceleration model to estimate free-fall dynamics [21]. The correspondence between the detected objects (localized objects) and tracks are found using a standard Hungarian algorithm [22].

Each object is assigned to a track, which consists of unique identity (ID), age, and position in the ROI. When the detected object is associated with a specific track, the tracker updates that track by Kalman filter correction. A track that exits the ROI is removed from the track list to reduce computational cost. Age is the number of frames after the track was first detected in the ROI; in our experiments, seeds stay within the ROI for about 12 ages during free-fall. With this flow, latency of both localization and tracking was $<$1.5 ms (Fig. 5 black line).

After the localization and tracking tasks are complete, single-seed images of each track are cropped from the ROI and used as the input of the classifier; single-seed images are cropped to 100×100 resolution and resized to 50×50 to reduce inference latency. Inference time of a CNN depends on several factors including input size, the number of weights, and GPU capability. CNNs with millions of weights require $>$5 ms of GPU inference time (based on GTX 1080Ti model). Therefore, such a CNN-based classifier cannot be directly applied to the 500 fps throughput applications. To solve this problem, we proposed the *batch inference only once* strategy that performs inference only once during free fall of seeds (Fig. 6). Because seeds for 12 ages on average, seeds can be captured about 12 times during free fall and the strategy performs inference once per object during object tracking. To accurately predict class by a single inference, the classifier must be trained on seeds in various poses (Section III-D).

The *batch inference only once* strategy proceeds as follows. At first, we defined *trigger age* as 5 by considering the inference time of the classifier. Seed tracks that have ages $<$5 have been assigned to a unassigned class (Fig. 6(a)). When one track becomes age 5, tracks with age $\leq$5 are bundled as batch images and are subject to batch inference (Fig. 6(b)). Because batch inference requires 5 to 8 ms, the CNN prediction is completed after approximately 3 to 4 frames have elapsed (1 frame $=$ 2 ms). After classification is complete, the predicted class is assigned to its corresponding track based on the ID information (Fig. 6(c), (d)). This procedure is repeated continuously (Fig. 6(e), (f)). Therefore, a class has four different states: (1) U: unassigned, in which objects are not subject to classification yet; (2) C: a state that requires 5–8 ms (3–4 frame ages), in which CNN is predicting classes of the given batch single-seed images; when the prediction is completed, the seed is assigned to either (3) SH (sunn hemp) or (4) W (weed). If *trigger age* is smaller than 5, the *batch inference only once* strategy becomes inefficient and causes some problems because GPU batch inference can require an additional inference before the previous inference has finished. In contrast, if *trigger age* is larger than 7, some seeds can exit the ROI before the inference has finished, and cannot be assigned to predicted class and sorted. Therefore, *trigger age* should be set to between 5 to 6 for stable operation of the *batch inference only once* strategy.

Using the proposed batch inference strategy, the proposed system with a CNN classifier achieved sorting at throughput of 500 fps. All single-seed tracks were evaluated by the CNN-based classifier as complete (Fig. 5).

C. Latency and Throughput

To be applicable to practical sorting, the proposed system must have high throughput; i.e., must recognize and sort a large number of seeds in a limited time. This requirement could be achieved when one camera can cover a wide horizontal range and analyze a large number of parallel seed streams. Estimated throughput was 8.5 kg/h when three seed streams were generated and covered by a camera (Fig. 5). Throughput can be increased by increasing the width of the ROI and the number of seed streams that it covers. Increase in the larger ROI would increase the inference latency that can be tolerated, and increase throughput, but the number of seeds in an image would increase; increased number of seeds in an image increases computational cost, so processing latency would increase. We therefore performed an experiment to find the maximum number of channels that our system with a single camera can process concurrently in $<$2 ms. We increased field of view and the number of channels that generate a single seed stream, and measured elapsed time

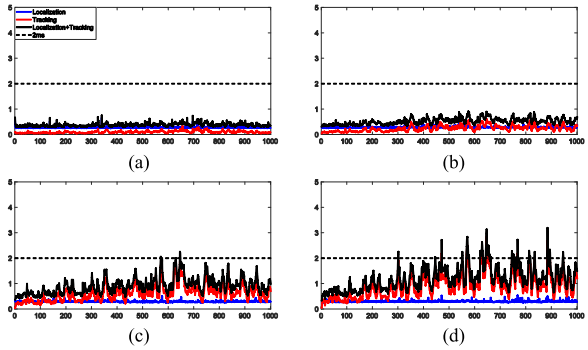

Fig. 7. Latency and throughput analysis. (a), (b), (c), and (d) latency of localization and tracking from different numbers of channels. Localization (blue) is not affected by the number of objects, but tracking time (red) increases according to the number of objects in ROI frame. Latency <2 ms of both localization and tracking is preserved until the number of channels becomes 24. (a) 6 channels (# of seeds: ∼9). (b) 12 channels (# of seeds: ∼16). (c) 24 channels (# of seeds: ∼29). (d) 30 channels (# of seeds: ∼36).

during processing in each experiment (Fig. 7). As the ROI width increased, the number of objects in an image frame increases. Increase in the number of objects had no effect on latency of image localization (Fig. 7 blue line), but increased the latency of the Kalman filter motion tracker (Fig. 7 red line). With 24 channels (eight times wider ROI than three channels), and the goal of 500 fps analysis, estimated throughput was 68 kg/h ($= 8 \times 8.5$). Throughput can also be increased by using additional cameras in parallel. If six cameras are used, as in BELTUZA (Section II-D), the estimated throughput becomes 408 kg/h, which is equivalent to commercial optical sorters.

D. Image Classifier

1) Model and Optimization: The GPU inference time is designed to be <8 ms while achieving high classification accuracy. To realize this goal, we trained the acquired seed data on several networks and compared their accuracy and latency. Also, input image was resized as 50 × 50 to reduce the latency; based on the results we selected ResNet-18 which has 11,190,082 weights and 5–8 ms latency as a classification network [15]. We used the categorical cross-entropy loss function

$$L(\hat{y}, y) = -\sum_i y_i \log \hat{y}_i, \tag{3}$$

where i is class index, y is target class distribution, and $\hat{y}$ is predicted class distribution. As an optimizer, Adam with learning rate `1e-3` was used [23].

2) Dataset Preparation: To satisfy the demand for super-high-purity and small sorting loss, the test accuracy of the classifier should be ≥99%. The core property of deep learning is that the performance of deep neural networks increases as the number of training data increases. To realize highly accurate classifier, three strategies have been used: acquisition of many data, data augmentation, and data balancing. In this study, acquisition of labelled data is relatively cheap because human-separated seeds have 100% purity, and the images of each known purified seed sample are easily acquired using our testbed. The seed images were acquired during free fall so the seed images have various rotational angles. Moreover, data augmentation including random rotation, shifting, and flipping was applied and it obviously contributed to the test accuracy improvement. We used binary classification because it was more accurate than five-categories classification. To train the network we used 132,428 sunn hemp seeds and 164,973 weed seeds. We used more weed seeds than sunn hemp seeds because the 'weed' class is composed of four species, so TN accuracy is degraded when the two classes have the same number of observations. To prevent model overfitting to sunn hemp class and improve the TN accuracy for realization of weed-free sorting, we manually balanced the data ratio and removed data bias.

TABLE II
TEST ACCURACY OF TRAINED CLASSIFIER

True	Inference	Correct	Wrong	Accuracy
Sunnhemp	Sunnhemp	10,165	28	99.725%
Cassia tora	Weed	9,989	28	99.721%
Cannabis	Weed	10,550	44	99.585%
Sesbania	Weed	10,452	14	99.866%
Morning glory	Weed	10,373	102	99.026%

TABLE III
TEST ACCURACY, PURITY, AND LOSS OF TRAINED CLASSIFIER

		Prediction			
		Positive	Negative	Purity	Loss
True	Positive	99.72%	0.28%	99.994%	1.495%
	Negative	0.46%	99.54%		
	Sorting	Accept	Reject		

3) Classification Accuracy: After training the network, we experimentally evaluated the test classification accuracy using another annotated set of seed data. The system achieved 99.72% TP and 99.54% TN, which correspond to 99.994% purity and 1.495% loss (Table. II–III). Although these are estimates and may not represent exact sorting accuracy, they are far superior to those of commercial sorting systems.

E. Seed Ejection based on Predictive Positions of Seeds

Solenoid valve is a mechanical system and it has lower bandwidth than that of a seed recognition. Even though we immediately operate the valve based on the processing result, the valve operation has some delay to shoot the target weeds. Therefore, the ejection strategy should consider the operation delay to exactly expel the target weeds. We considered sorting region and sorting timing. For the sorting region, the ROI frame is divided into three sorting regions in the horizontal direction and each sorting region is assigned to each of the three valves. The ejection region of each valve is placed under the ROI frame; three valves occupy each region. The solenoid valve (SMC corporation, VKF332-5G-M5) requires ∼10 ms to entirely open and close the valve after an operating signal arrives. The sorting timing is computed by Kalman filter prediction. The valves are placed under the ROI frame and constant acceleration dynamics can predict the distance that the seed has moved while valve is opening; in the absence of disturbance such as wind, the predicted position is reliable. Based on the valve position and valve

TABLE IV
COMPARISON OF *m*APS WITH THREE DIFFERENT DETECTORS

	Sunn hemp AP	Weed AP	*m*AP (IoU 0.4)
YOLO9000	86.36%	44.19%	65.29%
SSD	94.03%	87.32 %	90.68%
Ours (2,000)	**100.0**%	**100.0**%	**100.0**%
Ours (4,049)	**99.62**%	**99.62**%	**99.59**%

delay, we send the valve-operating signal when the target weed is in a vertical pixel position between 470 and 490, which corresponds to ages 10 to 12. results show the timing is matched (Multimedia).

IV. EXPERIMENTS

A. Comparison of Detection Precision with State-of-the-art Detectors

To evaluate the precision of the proposed object-detection method, we acquired verification data that include images of a mixture of falling seeds, and manually-labelled bounding boxes and classes. Using the newly labelled data, we compared the object-detection precision of the proposed system with state-of-the-art detectors: YOLO9000 and single shot multibox detector (SSD) [19], [20]. Both detectors have end-to-end architectures and they have to be trained on whole images (500×500 pixels) with their ground truth bounding boxes and classes. For the fair comparison, both detectors used same baseline network (ResNet-50) pre-trained on ImageNet [24]. We separated labelled data as training, validation, and test sets; 8,000 of training and 2,000 of validation data were manually labelled.

For the comparison, average precision (AP) and mean average precision (*m*AP) were used. The average precision is ratio of true positive among all predicted detection (summation of true positive and false positive) as

$$\mathrm{AP} = \frac{\text{Num. of TP}}{\text{Num. of TP} + \text{Num. of FP}}. \tag{4}$$

If intersection over union (IoU) is larger than threshold value, it is considered as detection (or hit); we used 0.4 IoU threshold. *m*AP is the mean value of average precision for each class ($m\mathrm{AP} = \sum_{c \in C} \mathrm{AP}_{c \in C} / N_C$ where $\mathrm{AP}_{c \in C}$ is AP for class c and N_C is the number of classes).

In the comparison results, the proposed system shows almost 100% of ideal precision (Table. IV); it shows all correct results for 2,000 validation data, but there are occasional failures when occlusion occurs in some test data (Multimedia). We think there are two reasons for this extremely high accuracy: 1) the network is well overfitted to this particular application, and 2) the training data is well balanced so that both TP and TN are accurate.

It should be noted that if we finely tuned SSD and YOLO9000 with many data, precision of them could be more improved; but we can argue that the proposed system that uses low-latency pipeline can have sufficiently high accuracy in this fixed environment.

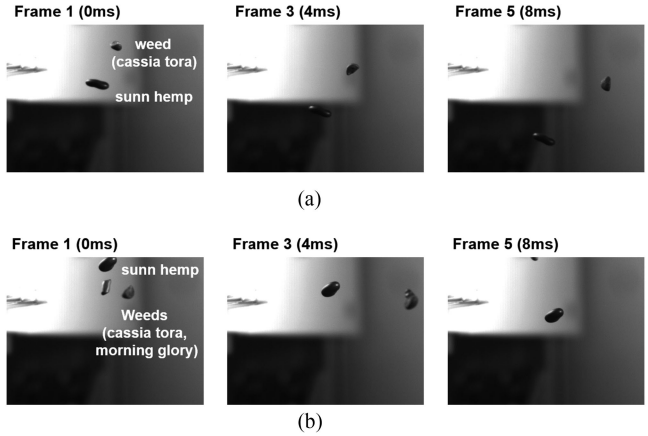

Fig. 8. Sequences of close-ups during sorting experiments. In both weed following sunn hemp (a) and weeds followed by sunn hemp (b) cases, air ejectors successfully blow the target weeds. All image sequences were taken at 500 fps.

B. Tracking Performance Evaluation

We should evaluate whether the tracking algorithm used in the proposed pipeline is accurately operated. Standard metrics to evaluate the tracking performance are multiple object tracking precision (MOTP) and multiple object tracking accuracy (MOTC) [25], [26]. For the 2,000 validation data, all assignments and IDs are correct (Multimedia); thus, both MOTA and MOTP are close to 1 and similar to detection results. We can also claim that the Kalman filter motion tracker has also sufficient performances to use in the fixed environment.

C. Sorting Experiment Results

We have shown that the proposed image-recognition system is highly accurate, but sorting accuracy is not guaranteed without real-time sorting experiment; we therefore performed sorting experiments using the proposed system to verify the feasibility. In actual experiments, there are factors that degrade sorting accuracy derived from hardware components such as inconsistent delay of valve, misalignment between valves and conveyor belt, and seed overlap; thus, sorting accuracy was not same with the accuracy of the image localization and tracking. We performed sorting experiments to evaluate that all these factors were appropriately set.

During the experiments, we recorded close-ups of the moment when seeds were ejected using additional high-speed camera (Multimedia). The close-ups show that weed seeds were removed well from the mixture of seeds in both weed seed following and followed by sunn hemp seed cases (Fig. 8). For quantitative analysis, we performed several experiments with a sample that contained known numbers of sunn hemp and weeds (Table V). In case 1, 1,000 sunn hemp seeds were tested to see how many sunn hemp seeds would be wrongly rejected and become losses. From case 2 to 6, the mixed seeds were used to analyze the sorting accuracy in various conditions. The result shows that in all cases there were one or zero fps in the accept box. Even though there were some FNs in the reject box, loss was extremely low due to low P_{sh} and many TPs. In conclusion,

TABLE V
QUANTITATIVE SORTING EXPERIMENT RESULTS

	Accept		Reject		Purity	Loss
	TP	FP	FN	TN		
Case 1	997	0	3	0	99.81%	0.47%
Case 2	10	0	0	3		
Case 3	19	1	1	3		
Case 4	10	0	0	1		
Case 5	16	1	1	2		
Case 6	13	0	0	5		
Total	1065	2	5	14		

the experimental results show the feasibility that the proposed image-recognition system can be applied to actual sorting system. Moreover, it is expected that a development of complete weed-free system through optimization of system alignment and improvement of hardware components can be made.

V. DISCUSSION AND CONCLUSION

In this letter, we developed an optical-seed-sorting system that achieves both high throughput and high purity by exploiting deep learning using an *inference only once* strategy. The *batch inference only once* method allows use of a CNN-based classifier, which has high accuracy but is not easy to apply in real-time systems because of its large inference latency. In quantitative evaluations our system achieved higher purity, classification accuracy and detection *m*AP, and lower loss and latency than commercial systems and state-of-the-art detectors did. The experiment results showed that the proposed image-recognition system can be directly applied to actual optical sorting systems. This system can be applied to acquire clean seed samples that are contaminated by noxious weeds.

The proposed system can also be used to sort other objects where results require high purity. Because this system exploits characteristics such as shape and texture to classify objects, it can detect inferior goods among a number of products. This extensibility to other industries is a useful feature of this system.

REFERENCES

[1] L. A. Herzenberg, D. Parks, B. Sahaf, O. Perez, M. Roederer, and L. A. Herzenberg, "The history and future of the fluorescence activated cell sorter and flow cytometry: A view from stanford," *Clinical Chem.*, vol. 48, no. 10, pp. 1819–1827, 2002.

[2] D. A. Hendrickson and A. Oberholster, "Review of the impact of mechanical harvesting and optical berry sorting on grape and wine composition," *Catalyst, Discovery Practice*, vol. 1, no. 1, pp. 21–26, 2017.

[3] M. Sofu, O. Er, M. Kayacan, and B. Cetişli, "Design of an automatic apple sorting system using machine vision," *Comput. Electron. Agriculture*, vol. 127, pp. 395–405, 2016.

[4] S. Jeong, Y.-M. Lee, and S. Lee, "Development of an automatic sorting system for fresh ginsengs by image processing techniques," *Human Centric Comput. Inform. Sci.*, vol. 7, no. 41, 2017.

[5] S. Hirano, N. Okawara, and S. Narazaki, "Near infra red detection of internally moldy nuts," *Biosci. Biotechnol. Biochem.*, vol. 62, no. 1, pp. 102–107, 1998.

[6] T. Pearson, "Machine vision system for automated detection of stained pistachio nuts," *LWT Food Sci. Technol.*, vol. 29, no. 3, pp. 203–209, 1996.

[7] A. Farsaie, W. McClure, and R. Monroe, "Design and development of an automatic electro-optical sorter for removing BGY fluorescent pistachio nuts," *Trans. ASAE*, vol. 24, no. 5, pp. 1372–1375, 1981.

[8] M. Pasikatan and F. Dowell, "Sorting systems based on optical methods for detecting and removing seeds infested internally by insects or fungi: A review," *J. Appl. Spectroscopy Rev.* vol. 36, pp. 399–416, 2001.

[9] A. Vibhute and S. Bodhe, "Applications of image processing in agriculture: a survey," *Int. J. Comput. Appli.*, vol. 52, no. 2, pp. 34–40, 2012.

[10] M. Bukovec, Ž. Špiclin, F. Pernuš, and B. Likar, "Automated visual inspection of imprinted pharmaceutical tablets," *Measur, Sci. Technol.*, vol. 18, no. 9, pp. 2921–2930, 2007.

[11] T. Pearson, "Hardware-based image processing for high-speed inspection of grains," *Comput. Electron. Agriculture*, vol. 69, no. 1, pp. 12–18, 2009.

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. 25th Int. Conf. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[13] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv:1409.1556, 2014.

[14] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2818–2826.

[15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[16] P. P. Rotar and R. J. Joy, "Tropic Sun' Sunn Hemp; Crotalaria juncea L." *Research extension series-College of Tropical Agriculture and Human Resources, University of Hawaii (USA)*, 1983.

[17] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Adv. Neural Inf. Process. Syst.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[18] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 779–788.

[19] J. Redmon and A. Farhadi, "Yolo9000: better, faster, stronger," arXiv:1612.08242.

[20] W. Liu *et al.*, "SSD: Single shot multibox detector," in *Eur. Conf. Comput. Vis*. Springer, 2016, pp. 21–37.

[21] N. Funk, "A study of the Kalman filter applied to visual tracking," *Project for CMPUT, University of Alberta*, Edmonton, AB, Canada, 2003.

[22] H. W. Kuhn, "The hungarian method for the assignment problem," *Naval Res. Logis.*, vol. 2, no. 1/2, pp. 83–97, 1955.

[23] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv:1412.6980, 2014.

[24] O. Russakovsky *et al.*, "Imagenet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, 2015.

[25] K. Bernardin and R. Stiefelhagen, "Evaluating multiple object tracking performance: the clear mot metrics," *J. Image Video Process.*, vol. 2008, 2008, Art. no. 1.

[26] A. Milan, L. Leal-Taixé, I. Reid, S. Roth, and K. Schindler, "Mot16: A benchmark for multi-object tracking," arXiv:1603.00831, 2016.