This CVPR 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.

GrainSpace: A Large-scale Dataset for Fine-grained and Domain-adaptive Recognition of Cereal Grains

Lei Fan[1,2] *† Yiwen Ding[1] * Dongdong Fan[1] Donglin Di[3] Maurice Pagnucco[2] Yang Song[2]
lei.fan1@unsw.edu.au {dingyiwen,fandongdong}@gaozhe.com.cn
didonglin@baidu.com morri@cse.unsw.edu.au yang.song1@unsw.edu.au
[1]Gaozhe Technology [2]The University of New South Wales [3]Baidu

Abstract

Cereal grains are a vital part of human diets and are important commodities for people's livelihood and international trade. Grain Appearance Inspection (GAI) serves as one of the crucial steps for the determination of grain quality and grain stratification for proper circulation, storage and food processing, etc. GAI is routinely performed manually by qualified inspectors with the aid of some hand tools. Automated GAI has the benefit of greatly assisting inspectors with their jobs but has been limited due to the lack of datasets and clear definitions of the tasks.

In this paper we formulate GAI as three ubiquitous computer vision tasks: fine-grained recognition, domain adaptation and out-of-distribution recognition. We present a large-scale and publicly available cereal grains dataset called ***GrainSpace****. Specifically, we construct three types of device prototypes for data acquisition, and a total of 5.25 million images determined by professional inspectors. The grain samples including wheat, maize and rice are collected from five countries and more than 30 regions. We also develop a comprehensive benchmark based on semi-supervised learning and self-supervised learning techniques. To the best of our knowledge, GrainSpace is the first publicly released dataset for cereal grain inspection,* `https://github.com/hellodfan/GrainSpace`.

1. Introduction

Cereal grains are the foundation of human civilization and are inextricably linked to our daily life. According to the data from the Food and Agriculture Organization of the United Nations in 2020 [1], the three types of cereal grains: wheat, maize and rice (see Figure 1), represent nearly 90% of the worldwide produce of cereal grains.

Grain determination is a crucial part in quality inspection and grade stratification, which provides guides and measures for grain circulation, storage, process and international trade. The majority of work for grain determination consists of chemical analysis and Grain Appearance Inspection (GAI). Chemical analysis is usually conducted by using various apparatus, but GAI still requires manual inspection with the aid of some hand tools ranging from sieves, dividers, to balances. In GAI, a batch of test samples are superficially inspected by professional inspectors in a kernel-by-kernel way. GAI can determine multiple metrics, such as impurities, damaged grains and cultivated varieties [21]. Taking GAI of wheat grains as an example, 60 grams (about 1600 grain kernels) are inspected and then divided into predefined groups manually, which requires 25 to 30 minutes for an inspector with 3 to 5 years' experience. It is thus highly desirable to develop an automated GAI.

Figure 1. Examples of wheat, maize and rice grain kernels.

Over the past few years, deep learning techniques have achieved remarkable success in many computer vision applications, such as recognition (ImageNet [37]), detection (MS-COCO [26]), segmentation (Cityscapes [9]) and video understanding (YouTube-8M [2]). There are however two main challenges to apply deep learning models to GAI. First, in-depth domain knowledge of GAI is required in order to formulate the grain determination problem into proper computer vision tasks. Second, developing deep learning-based methods for GAI requires high-quality datasets covering a comprehensive representation of

*Equal contribution.
†Work done when interning at Gaozhe technology.

the large variety of cereal grains.

In our work, we perform in-depth analysis of the characteristics of cereal grain and consider the real-world requirements of GAI. We formulate GAI into three fundamental computer vision tasks: fine-grained recognition, domain adaptation and out-of-distribution recognition. We built three kinds of device prototypes that can capture images of cereal grains efficiently. Then, we construct a large-scale dataset containing 5.25 million images concerning three types of cereal grains (wheat, maize and rice) collected from five countries and more than 30 regions. The raw grain samples were processed manually by nine inspectors over more than 4 years. Furthermore, we develop a benchmark on our proposed dataset by employing advanced techniques like semi-supervised learning and self-supervised learning to address the challenges in fine-grained recognition, domain adaptation and out-of-distribution recognition. Our experimental results show that the developed approaches can obtain substantial improvements and make automated GAI feasible. Our contributions are summarized as follows:

- A large-scale and publicly available cereal grain dataset called ***GrainSpace***, containing 5.25 million images of wheat, maize and rice grains, is constructed.
- Based on our in-depth analysis of GAI, we formulate GAI-related work into three computer vision tasks, including fine-grained recognition, domain adaptation and out-of-distribution recognition.
- An initial benchmark is developed to address the above tasks and we demonstrate promising performance on the *GrainSpace*.

2. Related Work

GAI provides a foremost assessment on the quality of grains, assisting grading, cleaning and separation of grains. As the appearance and physical characteristics of grains are highly variable, GAI is error-prone even for trained inspectors. There are high demands in automated GAI that have the benefit of greatly assisting inspectors. However, there are two main challenges in building automated GAI: what GAI-related tasks we should focus on and how to construct a high-quality cereal grain dataset.

GAI-related work: In general, GAI is used for providing accurate classification and identification of various grains [45]. Limited by sensor technologies and computational resources, early studies [44, 3] employed machine vision to classify five types of wheat (barley, oats, rye, wheat and durum wheat) or impurities (stones, soil and weeds) based on statistical information, such as color, morphological or textural variations. Some researchers utilized neural networks to identify the varieties of rice and wheat. For example, Zapotoczny [48] and Golpour *et al*. [14] analyzed textures of grains to classify 11 varieties of spring/winter wheat and 5 brown/white rice cultivars. Guzman *et al*. [15] and Shantaiya *et al*. [39] developed algorithms to identify five groups of rice in Philippines and six varieties of rice seeds in Chhattisgarh, respectively. In this paper we comprehensively analyze GAI-related tasks, such as identifying the damaged and unsound grains that include grains damaged by pressure, pests and fungus, and then we formulate GAI into three computer vision tasks: fine-grained recognition, domain adaptation and out-of-distribution recognition.

Cereal grain dataset: Advances of deep learning have revolutionized multiple real-world fields such as medical analysis [42], autonomous driving [40] and agriculture [24]. The success of deep learning is mainly attributed to abundant computational resources, well-designed network architectures and large-scale datasets. In particular, high-quality datasets, such as ImageNet [37], Pascal VOC [11], MS-COCO [26], Cityscapes [9] and Kinetics [25], are essential for many computer vision tasks, *e.g*., image classification [18], object detection [35], semantic segmentation [29] and video understanding [12]. For the last several years, many researchers have also investigated more industry-related visual tasks, such as anomaly detection [5], sewer detection [16], food recognition [31] and nutrition estimation [43]. However, to the best of our knowledge, there are few publicly available cereal grain datasets. Most of the previous studies [46, 34, 33] focus on building devices for image acquisition with specific sensors, such as hyperspectral imaging. In this work we built three kinds of device prototypes: P600, G600 and M600. P600 and G600 consist of industrial cameras, grain holding platforms and lighting sources for illumination. M600 is based on a smartphone that is low-cost and ideal for widespread deployment. We create a total of 5.25 million images that the raw grain samples are from multiple countries and regions and are manually pre-processed by nine trained inspectors carefully.

3. *GrainSpace*

In this section, we present GAI as three challenges related to computer vision tasks (see Figure 2), and describe device prototypes along with procedures for data processing (see Figure 4) and data distribution. Note that more detailed descriptions are included in the supplementary material.

3.1. Challenges

Over recent decades, GAI as a conventional but crucial part of grain determination is routinely performed manually. Each grain kernel in a batch of grain samples is inspected carefully. The main inspection work focuses on determining whether the grain kernel is Damaged and Unsound (DU), and identifying the sub-type of grain kernels.

In accordance with ISO5527-Cereals [21], wheat grains can be categorized as NORMAL and six types of DU grains: FUSARIUM & SHRIVELLED (F&S) grain, SPROUTED

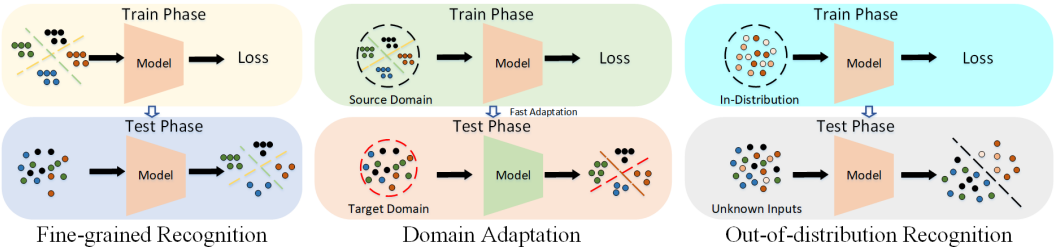


Figure 2. Illustrations of three GAI-related challenges: fine-grained recognition, domain adaptation and out-of-distribution recognition.

Table 1. Examples of normal and DU wheat grains.

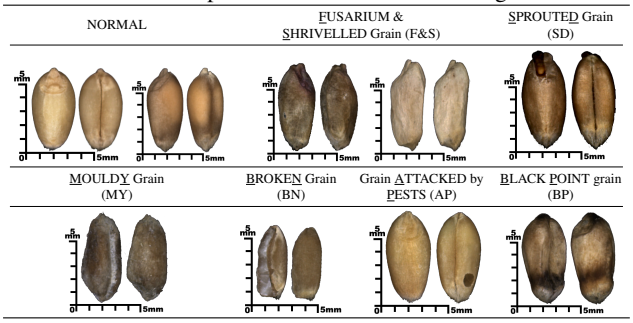
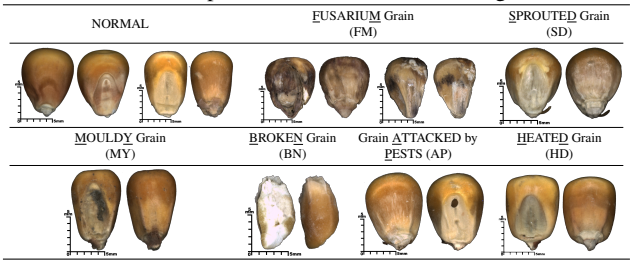


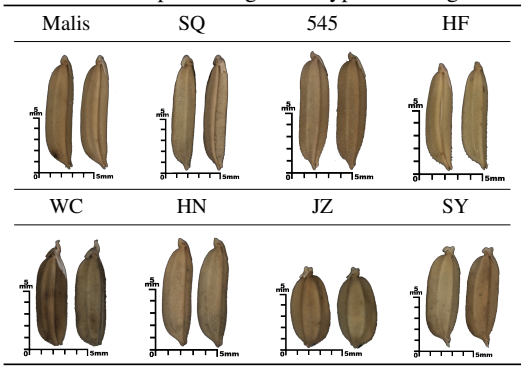
Table 2. Examples of normal and DU maize grains.



(SD) grain, MOULDY (MY) grain, BROKEN (BN) grain, grain ATTACKED by PESTS (AP) and BLACK POINT (BP) grain (see Table 1). Maize grains are also grouped into NORMAL and six DU-grain types: FUSARIUM (FM) grain, SD grain, MY grain, BN grain, AP grain and HEATED (HD) grain (see Table 2). Among these grains, F&S, FM, MY and BP grains indicate the proportion of grains that are contaminated by fusarium or fungus, etc; SD, AP and HD grains correspond to the nutrient content of grains. In terms of rice grains, Table 3 illustrates 8 sub-types, in which Malis, SQ and 545 belong to "Thai Hom Mali Rice" are 2 to 4 times more expensive than the other kinds of rice grains. Different sub-types of rice grains look very similar but these rice grains may have large gaps in nutrient content, taste and the most important part: price. Therefore, it is an important GAI task to identify the sub-types of rice grains, especially for some rare sub-types.

While sub-type identification is naturally a classification problem, based on our experimental studies, we discovered that there are more challenges associated with this task. In

Table 3. Examples of eight sub-types of rice grains.



particular, we need to solve the fine-grained recognition, domain adaptation and out-of-distribution recognition problems (see Figure 2).

Fine-grained recognition: Grains of the same species usually share similar appearance characteristics in terms of shape, color and texture. However, there are some tiny yet crucial differences between normal and DU grain kernels, and between different sub-types. For example, the tiny pest hole in a wheat grain is only of $1 \times 1mm^2$ (see Table 1). In order to effectively differentiate the subtle differences, this thus becomes a Fine-Grained Visual Categorization (FGVC) problem. FGVC has typically been applied to distinguish bird species [4] and car models [47], etc. Similarly, we formulate DU-grain and sub-types identification as FGVC tasks.

Domain adaptation: Usually, due to geographical and climate reasons, different countries or regions have distinct differences in the varieties of grain. These differences not only exhibit in the shape and size of grain kernels, but also show in the texture and color distribution. Table 1 illustrates two examples of normal wheat grains with different colors. Despite these differences, qualified inspectors can still obtain correct results because the prominent characteristics of grains are clearly discriminated. This requirement is coherent with Domain Adaptation (DA). The objective of DA is to enhance the performance on the target domain based on the model that is trained with the existing source domain. Taking DU-grain identification as an exam-

ple, in most cases, only grain samples from some regions (source domain) can be obtained, and the model trained on the source may be tested on some grain samples from unknown regions (target domain). In addition, since we build and employ different device prototypes to acquire data, the data across different prototypes can also be seen as different domains.

Out-of-distribution recognition: One of the crucial but difficult GAI tasks is to identify the proportion of some specified sub-types of grains. Most of the time, food factories or storage facilities require only specific sub-types grains (*e.g.*, "Thai Hom Mali Rice": Malis, SQ and 545), but the test samples may have other sub-types of grains. We consider that such requirement is related to out-of-distribution (OOD) recognition. OOD, containing anomaly detection, aims at identifying whether the input belongs to the in-distribution (of interest) or not (out-of-distribution). The expected sub-types of grains can be regarded as in-distribution, and all other kinds of grains will be seen as out-of-distribution. Similarly, DU-grain evaluation also can be considered as OOD recognition. Note that, in comparison with common OOD tasks, OOD recognition related to GAI is mixed with fine-grained recognition and is more challenging because the differences between in-distribution and out-of-distribution data are minor.

3.2. Data Acquisition

To construct the cereal grains dataset, devices for data collection are prerequisites. We intend to design devices to capture accurate and realistic photographs of grain kernels. However, there are two challenges in capturing high-quality images of grain kernels: 1) To capture overall appearance information of grain kernels, dual or multiple cameras should be set at appropriate angles around grain kernels. 2) Compared to natural objects (dog or building, etc.), grain kernels with tiny sizes (usually smaller than $8 \times 8 \times 4mm^3$) impose huge difficulties on the environment including stability and lighting condition, etc.

Prototypes: We build three kinds of device prototypes: Professional-600 (P600), General-600 (G600) and Mobile-600 (M600) (see Figure 3.a). Specifically, P600 mainly consists of dual industrial cameras with light sources and a conveyor belt for automatically feeding grain kernels, G600 consists of an industrial camera with light sources and a conveyor belt, and M600 consists of a mobile phone and a holder for fixing the phone. We design a robotic automation mechanism to manipulate P600 and G600 to implement data sampling automatically with higher sampling efficiency but also higher complexity, while M600 requires placing grain kernels manually. Among these devices, P600 with dual cameras is able to capture a larger Effective Receptive Field (ERF) but the manufacturing cost is very high, whereas G600 and M600 with one camera could only capture a single view photograph of grain kernels under a moderate ERF. We compare these device prototypes in terms of cost, ERF, reproducibility, automaticity and complexity in Figure 3.b.

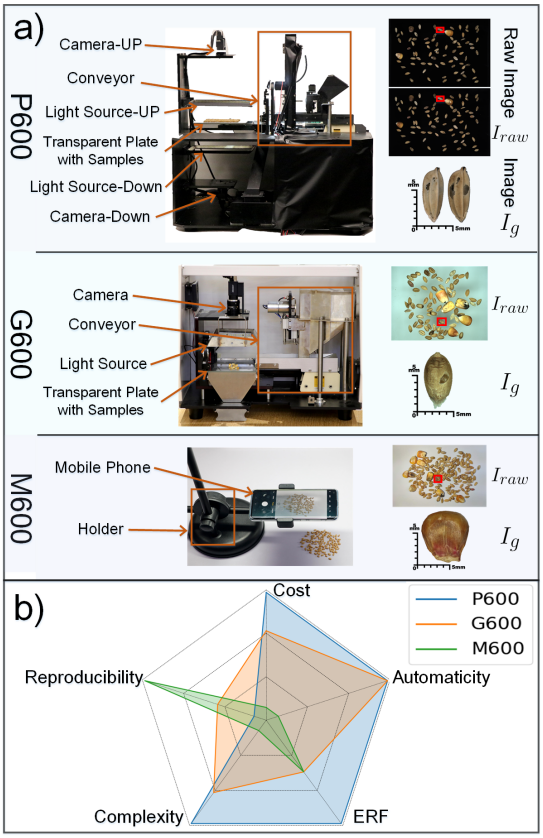


Figure 3. **a)** The prototypes and captured photographs of P600, G600 and M600; **b)** The radar diagram of performance comparisons among these prototypes.

Data Processing: Our goal is to construct a high-quality cereal grains dataset. However, if we attempt to collect grain images in a kernel-by-kernel way, it is extremely time-consuming and infeasible to be applied in the real world. Therefore, to obtain data efficiently, we establish a data processing procedure based on our prototypes (see Figure 4). Specifically, following the ISO24333-Cereal Sampling [20], various impurities (extraneous and inorganic matter, etc.) and foreign cereals are carefully picked out from raw grain samples (obtained from granaries or freighters) by inspectors with tweezers and sieves. Then, grain samples without impurities are manually divided into several groups in accordance with predefined categories. For each specific category L, samples are sent to devices in batches to obtain N raw images $\{I^1_{raw}, \ldots, I^N_{raw}\}$, in which each I_{raw} contains many grain kernels that share the same label of category L. Single-kernel images I_g are then cropped from

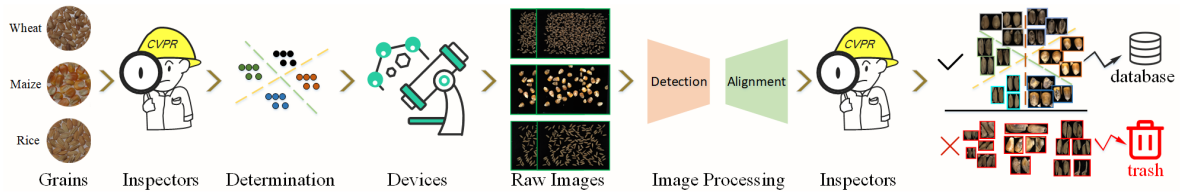


Figure 4. Overview of data acquisition. Grain kernels are determined and divided into predefined categories. Grain kernels from a group share the same category L and are delivered into devices to obtain raw images I_{raw}. I_{raw} with many kernels is processed to generate many kernel-wise images I_g via detection and alignment stages. Finally, inspectors filter out low-quality images.

I_{raw} via detection and alignment stages where a rotation-invariant object detector based on YOLOv5 [23] is introduced to localize all grain kernels with various orientations. All I_g are paired with the original category L as the ground truth. It is worth noting that some I_g captured by G600 or M600 may not display prominent features due to the single-camera view, but we still keep these images with original label L because we expect to explore the limitation of advanced computer vision methods.

3.3. Data Distribution

All grain samples were collected from 5 countries and more than 30 regions during the period of 2017-2021 (details in supplementary material). Table 4 shows detailed information in terms of category, region, weight and the number of grain kernels for each species of grain. Among these samples, wheat grain samples (near 150 kilograms, 4.1 million grain kernels) are obtained from 50 tons of wheat, in which 1.6 million grain kernels are divided into 7 categories manually and 2.5 million grain kernels without labels are used for exploring unsupervised methods. Similarly, maize grain samples (near 95 kilograms, 0.3 million grain kernels) are obtained from 50 tons of maize, in which 0.16 million grain kernels are grouped into 7 classes and 0.14 million grain kernels without labels are also employed for unsupervised methods. Rice grain samples (near 22 kilograms, 0.82 million grain kernels) are obtained from 0.8 tons of rice (8 sub-types of rice, each of which is 100 kilograms).

Table 4. Information of raw wheat, maize and rice grains.

Species	Category	Region	Num. Grain Kernels	Weight
Wheat	7	22	4,129k	150 kg
Maize	7	8	299k	95 kg
Rice	8	8	820k	22 kg

Overall, *GrainSpace* contains a total of 5.25 million images, and the distribution is shown in Figure 5. To avoid potential ethical issues or privacy restrictions, we erased real source information and adopted R_N as substitutions for data anonymization. Wheat and maize images are divided into labeled and unlabeled groups corresponding to raw inspected and un-inspected grain kernels, respectively. Note that all grain kernels (including unlabeled kernels) are pre-processed (*e.g.*, to remove impurities) by inspectors manually, and labeled kernels are further determined and classified into predefined categories.

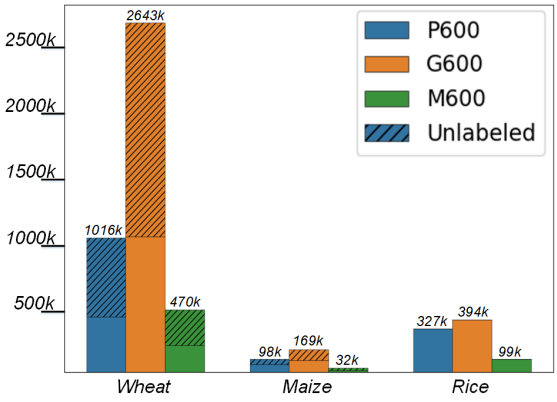


Figure 5. The distribution of *GrainSpace*.

Wheat: All wheat grain kernels sampled from 22 regions are divided into 3 groups according to region information, and a total of 4,129k images including 1,638k and 2,491k of labeled and unlabeled images respectively. In fact, since the real percentages of Damage and Unsound (DU) wheat grains account for less than 2% in raw wheat grains, gathering a large number of DU wheat grains is tremendously labor-intensive and costly. To keep a balance of data distribution, we tried our best and collected a total of 111k, 180.5k and 26.5k images of DU wheat grains by using P600, G600 and M600 respectively (see Table 5).

Table 5. Detailed statistics of wheat grain images.

Region	Device	NORMAL	Damaged and Unsound Wheat Grains						
			F&S	SD	MY	AP	BN	BP	Total
R_{1-14}	P600	216k	3.4k	3.4k	3.4k	3.4k	3.4k	3.4k	20.4k
	G600	756k	12k	12k	12k	12k	12k	12k	72k
	M600	127k	1.7k	1.7k	1.7k	1.7k	1.7k	1.7k	10.2k
R_{15-18}	P600	40k	0.8k	36k	1.8k	1.2k	5k	4.2k	49k
	G600	40k	0.8k	36k	5.5k	3.5k	5k	4.2k	55k
	M600	28k	0.6k	6k	1k	1k	2k	0.4k	11k
R_{19-22}	P600	49k	0.6k	27k	0.6k	0.8k	5.2k	7.4k	41.6k
	G600	47k	0.6k	36k	1.8k	2.5k	5.2k	7.4k	53.5k
	M600	18k	0.6k	2k	0.3k	0.7k	1k	0.7k	5.3k

Maize: All maize grain kernels are sampled from 8 regions, and a total of 299k images containing 159k and 140k of labeled and unlabeled images, respectively. Considering the scarcity of DU maize grain kernels similar to wheat samples, we tried our best and collected a total of 38k, 49.4k and 8.6k DU maize grain images by using P600, G600 and M600 respectively (see Table 6).

Table 6. Detailed statistics of maize grain images.

Device	NORMAL	Damaged and Unsound Maize Grains						
		FM	SD	MY	AP	BN	HD	Total
P600	20k	9k	3.4k	5k	9k	7k	4.6k	38k
G600	40k	10k	4k	10k	10k	10k	5.4k	49.4k
M600	4k	1k	0.4k	3k	1k	2k	1.2k	8.6k

Rice: In distinction to wheat and maize, the main challenge related to rice is to recognize the sub-type of test samples. We collected 8 sub-types of rice grain kernels from 8 regions respectively, and a total of 820k images consisting of 327k, 394k and 99k images captured by P600, G600 and M600 respectively (see Table 7).

Table 7. Detailed statistics of rice grain images.

Device	Categories of Rice Grains							
	Malis	SQ	545	HF	WC	HN	JZ	SY
P600	62k	30k	80k	40k	17k	40k	18k	40k
G600	80k	40k	80k	80k	16k	40k	18k	40k
M600	12k	8k	13k	14k	13k	13k	13k	13k

4. Benchmark

In this section, we present a comprehensive evaluation of advanced computer vision techniques as an initial benchmark for future work on *GrainSpace*. For these GAI-related challenges, we employ several classical and state-of-the-art methods and introduce semi-supervised and self-supervised learning techniques. Note that more detailed results are included in the supplementary material.

4.1. Experimental Setting

In all experiments, we randomly split each type of data into 80%, 10% and 10% of training, validation and test sets. We adopt PyTorch [32] as our experiment framework based on a GPU platform with 8 $\times$ Nvidia RTX 2080Ti. In order to keep fair comparison, all models are trained from scratch without pretraining on other datasets (*e.g.*, ImageNet [37]). Since the data distribution is heavily imbalanced, both precision and recall cannot appropriately reflect the performances of models. Therefore, we select the Macro F1-score as experimental measurement. Taking fine-grained recognition of wheat with N class as an example, we calculate N F1-score for each category, and the overall F1-score is obtained by averaging these F1-scores ($\frac{1}{N}\sum_n^N(F1_n)$). This section only reports the Macro F1-score and more detailed information are included in supplementary material.

4.2. Fine-grained Recognition

Considering that wheat data captured by different prototypes are divided into three region groups, we conducted 27 experiments based on ResNet50 (R50) [18], DCL [8] and Swin Transformer (SwinT) [27] (see Table 8). Among these methods, R50 is one of the most classical models, DCL is an advanced fine-grained recognition method, and SwinT is based on the popular transformer technique.

Table 8. Performance of R50, DCL and SwinT on wheat data: regions vs. device prototypes.

Model	R_{1-14}			R_{15-18}			R_{19-22}		
	P600	G600	M600	P600	G600	M600	P600	G600	M600
R50 [18]	**93.9%**	**80.1%**	87.6%	80.0%	76.5%	**79.7%**	70.1%	**76.1%**	**76.1%**
DCL [8]	92.5%	79.1%	**87.9%**	**82.1%**	**77.2%**	76.1%	**73.9%**	74.9%	72.4%
SwinT [27]	56.5%	39.2%	64.0%	49.8%	58.5%	43.9%	44.0%	51.3%	53.4%

We observe that R50 and DCL (R50 backbone) obtain all-sided advantages in all regions and prototypes, whereas SwinT has collapse performance on R_{1-14} (G600), R_{15-18} (P600 and M600), etc. The unsatisfactory results show the potential challenges of *GrainSpace* that require higher ability of models' generalization and adaptation. Figure 6 shows some visualization examples based on CAM technique [49] with DCL [8] models. To simplify experiment settings and save computational resources, follow-up experiments are mainly based on R50 as the backbone.

Next, we conducted 15 experiments on wheat, maize and rice data without considering region information (see Table 9), in which 6 experiments are conducted with a combination of G600 and M600 data, since these data are captured via a single camera. We observe that performance of wheat experiments are moderate but maize and rice obtain good results, which means wheat data from different regions should be processed carefully. With a package of G600 and M600 data, the performance from M600 data are heavily degraded, which, we consider, is mainly due to the imbalanced data distribution between G600 and M600. In addition, we utilize unlabeled data by introducing semi-supervised learning (MixMatch [6]) to wheat and maize experiments. All wheat experiments gain significant improvements but maize group has a little decrease. We consider that the different results are due to the ratio between labeled and unlabeled data (wheat 1:1.52, maize 1:0.88), and the smaller volume of unlabeled maize data should be used in more elaborate ways.

We further introduce self-supervised learning to explore unlabeled data, and apply MoCo [17] that is a powerful framework based on contrastive learning. We conducted 45 experiments on wheat, maize and rice data without considering region information (see Table 10). Following a common evaluation protocol [17, 7], we evaluated the perfor-

Table 9. Performance of device prototypes on wheat, maize and rice data. (+ and - denote results obtained from MixMatch [6]).

Species	Training set			Test set		
	P600	G600	M600	P600	G600	M600
Wheat	✓			68.5%+10.7%	-	-
		✓		-	63.5%+5.2%	-
			✓	-	-	59.4%+10.7%
		✓	✓	-	63.4%+4.5%	14.8%+14.7%
Maize	✓			94.0%-2.6%	-	-
		✓		-	86.6%-2.2%	-
			✓	-	-	82.8%-6.4%
		✓	✓	-	85.3%-1.6%	33.8%+24.3%
Rice	✓			99.2%	-	-
		✓		-	98.9%	-
			✓	-	-	93.0%
		✓	✓	-	98.7%	26.8%

mance by linear probe on the frozen features extracted from pretrained models, in which a supervised linear classier is trained with different proportions of unlabeled data. Almost all experiments show that a large proportion of unlabeled data and few labeled data can obtain comparable performance, which verifies that self-supervised learning has high potential in these tasks.

Table 10. MoCo [17] performance of device prototypes on wheat, maize and rice data.

Species	Training set			Test set	Labeled data proportion		
	P600	G600	M600		1%	10%	100%
Wheat	✓			P600	57.4%	**60.0%**	56.7%
		✓		G600	**65.3%**	63.4%	61.9%
			✓	M600	31.6%	**45.6%**	45.5%
		✓	✓	G600	58.2%	**60.2%**	59.6%
		✓	✓	M600	37.3%	**41.1%**	38.7%
Maize	✓			P600	17.2%	52.7%	**72.4%**
		✓		G600	12.3%	52.4%	**61.9%**
			✓	M600	6.9%	10.5%	**38.7%**
		✓	✓	G600	19.1%	54.3%	**62.8%**
		✓	✓	M600	9.7%	44.1%	**51.3%**
Rice	✓			P600	10.3%	44.2%	**49.0%**
		✓		G600	34.2%	54.5%	**70.4%**
			✓	M600	10.6%	16.2%	**32.1%**
		✓	✓	G600	37.9%	50.0%	**76.8%**
		✓	✓	M600	11.4%	44.2%	**50.2%**

4.3. Domain Adaptation

In *GrainSpace*, different regions of wheat data have diverse appearance although DU grains share common features, and thus different regions can be regarded as different domains. We evaluate domain adaptation (DA) performance by adopting three classical and advanced methods: CDAN [30], MCD [38] and MCC [22]. Among these methods, CDAN incorporates two conditioning strategies for guaranteeing model's discriminability and transferability, MCD attempts to align distributions of source and target by maximizing the output discrepancy between two classifiers, and MCC tries to minimize the class confusion between the correct and ambiguous classes for target examples. Since the appearance of wheat data vary across regions and device prototypes, we comprehensively conducted 72 experiments in terms of each combination of region and prototype (see Table 11). Almost all experiments obtain dramatic decreases, which may be attributed to these DA methods that are designed for common objects (*e.g.*, buildings) among different domains. However, in comparison with natural images with salient objects, the differences of wheat grains among different regions are minor but prominent. We regard that a possible solution is to enforce model to focus on local information based on existing DA techniques.

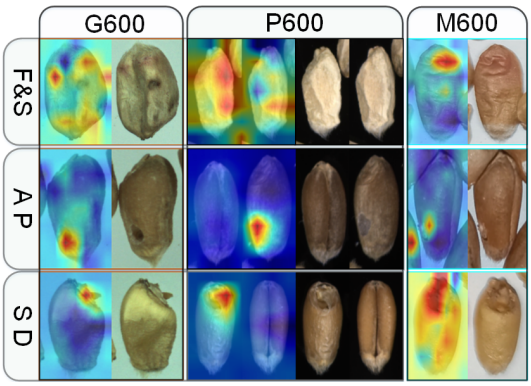


Figure 6. Examples of CAM-based visualization (DCL [8]).

Table 11. Performance of DA methods on wheat data: regions vs. device prototypes (order by P600, G600, M600).

Method	$R_{1-14} \rightarrow R_{15-18}$	$R_{15-18} \rightarrow R_{19-22}$	$R_{19-22} \rightarrow R_{1-14}$
Source Only	42.9%, 18.9%, 22.7%	52.9%, 16.1%, 46.2%	26.1%, 33.4%, 21.3%
CDAN [30]	-15.4%, -1.6%, -8.4%	-9.2%, +7.1%, -3.6%	+8.3%, -9.5%, +12.6%
MCD [38]	-22.9%, -8.6%, -10.8%	-15.2%, +8.3%, -18.3%	+0.9%, -12.3%, -1.6%
MCC [22]	-11.1%, +1.9%, -7.2%	-12.8%, +3.4%, -17.3%	-0.5%, -12.4%, -0.3%
Method	$R_{15-18} \rightarrow R_{1-14}$	$R_{19-22} \rightarrow R_{15-18}$	$R_{1-14} \rightarrow R_{19-22}$
Source Only	17.6%, 16.2%, 22.6%	45.6%, 26.6%, 48.2%	46.7%, 28.5%, 26.6%
CDAN [30]	+14.0%, +1.1%, +4.2%	-4.7%, -4.9%, -8.8%	-13.5%, -12.5%, -10.2%
MCD [38]	+9.6%, +4.7%, -4.5%	-9.2%, +3.2%, -29.2%	-25.7%, -13.7%, -13.8%
MCC [22]	+10.4%, -0.8%, -2.0%	-12.5%, -3.0%, -20.4%	-13.8%, -8.9%, -10.4%

Moreover, we conducted another 72 DA experiments on all wheat, maize and rice data by treating the device prototypes as different domains without considering region information (see Table 12). We observe that the majority of DA experiments obtained large improvements in comparison with source only experiments, which verifies that data from different device prototypes have potential to be used collectively to achieve high performance. It is obvious that the results adapting between G600 and M600 decrease heavily on the wheat data, and we are still analyzing the underlying reasons.

4.4. Out-of-distribution Recognition

In some cases, only several specific sub-types of rice grains are accepted and purchased by food factories or

Table 12. DA method performance of device prototypes on all grain data (order by wheat, maize, rice).

Method	P600→G600	G600→M600	M600→P600
Source Only	11.6%, 21.5%, 8.7%	13.2%, 29.9%, 23.1%	6.6%, 13.2%, 4.5%
CDAN [30]	+6.9%, +3.8%, +31.0%	+0.2%, -9.2%, +5.0%	+5.5%, +8.8%, +10.8%
MCD [38]	+2.9%, +5.8%, +4.9%	+2.4%, +0.2%, -13.8%	+6.1%, +8.8%, +9.4%
MCC [22]	+0.8%, +5.4%, +22.1%	+0.2%, -1.0%, +7.1%	+5.5%, +4.1%, +11.6%
Method	G600→P600	M600→G600	P600→M600
Source Only	12.1%, 7.6%, 27.1%	25.7%, 21.7%, 56.5%	4.4%, 17.9%, 11.3%
CDAN [30]	+0.9%, +25.2%, +8.3%	-10.2%, -2.2%, -15.0%	+11.3%, +0.1%, +8.8%
MCD [38]	+0.1%, +27.5%, -16.4%	-9.5%, +4.2%, -46.2%	+10.8%, +1.0%, -1.7%
MCC [22]	+0.4%, +18.2%, -3.5%	-7.4%, -0.5%, -16.5%	+5.2%, -4.0%, +12.6%

traders, and recognizing these grains can be treated as out-of-distribution (OOD) recognition. We combine data of specific categories to create one-class dataset configurations, and train OOD models on one class by employing three advanced methods: Deep SVDD [36], Rot [19] and CSI [41]. Specifically, Deep SVDD trains a model by minimizing the volume of a hypersphere that encloses data representations, Rot utilizes self-supervision to boost the identification on near-distribution outliers, and CSI introduces contrasitve learning into OOD problems to learn better visual representations. Following previous studies [10, 19, 41], the area under the receiver operating characteristic curve (AUROC) is employed to evaluate OOD models. A larger value of AUROC means better performance, and a value of 50% means random guess.

For P600 rice data (results for G600 and M600 are included in the supplementary material), we set 9 OOD experiments with three kinds of data configurations (see Table 13) where (Malis, SQ, 545) belong to "Thai Hom Mali Rice", (HF, WC, HN) share the similar price, and (JZ, SY) are sampled from the same province. We obverse that each OOD method performed moderate results on several data combinations but all experimental results are less that 80%, which means there are a large room for further exploration.

Table 13. OOD method performance on P600 rice data (✓ denotes this group is in-distribution).

Method	Malis	SQ	545	HF	WC	HN	JZ	SY	AUROC
Deep SVDD [36]	✓	✓	✓						62.5%
				✓	✓	✓			46.5%
							✓	✓	62.7%
Rot [19]	✓	✓	✓						61.1%
				✓	✓	✓			**64.1%**
							✓	✓	57.5%
CSI [41]	✓	✓	✓						**70.9%**
				✓	✓	✓			50.8%
							✓	✓	**77.3%**

In addition, the identification of DU grains also can be considered as OOD recognition. We conducted 12 OOD experiments on P600 wheat and maize data (see Table 14, G600 and M600 experiments are included in the supplementary material), in which (F&S, MY, BP) or (FM, MY, HD) are grouped together because that these kinds of DU grains have deleterious effects on health. In these evaluations, Rot and CSI achieve the highest performance of 68.5% and 71.6% on the "deleterious effect" group of wheat and maize respectively, and these performance are comparable and prove that treating DU-grain recognition as an OOD recognition is feasible and more suitable for applying in real-world applications.

Table 14. OOD method performance on P600 wheat and maize data (✓ denotes this group is in-distribution).

Species	Method	Normal	F&S	SD	MY	AP	BN	BP	AUROC
Wheat	Deep SVDD [36]	✓		✓		✓	✓		53.1%
			✓		✓			✓	56.0%
	Rot [19]	✓		✓		✓	✓		66.4%
			✓		✓			✓	**68.5%**
	CSI [41]	✓		✓		✓	✓		**70.3%**
			✓		✓			✓	60.2%
Species	Method	Normal	FM	SD	MY	AP	BN	HD	AUROC
Maize	Deep SVDD [36]	✓		✓		✓	✓		**69.2%**
			✓		✓			✓	43.1%
	Rot [19]	✓		✓		✓	✓		66.2%
			✓		✓			✓	67.8%
	CSI [41]	✓		✓		✓	✓		60.5%
			✓		✓			✓	**71.6%**

5. Conclusions and Future Work

In our study, we conducted an in-depth analysis of GAI and formulated GAI into three common computer vision tasks: fine-grained recognition, domain adaptation and out-of-distribution recognition. We created a publicly available large-scale grain cereal dataset: *GrainSpace*. For data acquisition, we built three kinds of device prototypes and established a comprehensive data processing procedure. Then, we collected a total of 5.25 million grain kernels images containing 4129k, 299k and 820k images of wheat, maize and rice, respectively. The raw grain kernels in *GrainSpace* were sampled from five countries and more than 30 regions across four years. In addition, we developed a benchmark on *GrainSpace* with comprehensive experimental analysis. We observed substantial improvements by introducing advanced computer vision techniques such as semi-supervised learning and self-supervised learning.

The main challenge in GAI is to identify the minor differences among different grain kernels. Due to the variety and diversity of grain kernels, models should be generalizable and adaptive for both existing data and unknown grain kernels. On the one hand, the number of UD grains is far lower than normal grains, which can be seen as a natural long tailed classification problem [28]. In addition, in our current work, the impurities, extra matters and foreign cereals are removed manually, which could be automated using computer vision techniques like open-set detection [13] etc. We hope that *GrainSpace* can stimulate and draw more attention to the development of intelligent agriculture, and we believe computer vision techniques can revolutionize GAI-related applications.

References

[1] World Food Situation, url = https://www.fao.org/worldfoodsituation/csdb/en, time = 2021-10-7.

[2] Sami Abu-El-Haija, Nisarg Kothari, Joonseok Lee, et al. Youtube-8M: A large-scale video classification benchmark. *arXiv preprint arXiv:1609.08675*, 2016.

[3] Basavaraj S Anami and D Savakar. Effect of foreign bodies on recognition and classification of bulk food grains image samples. *J. Appl. Comput. Sci*, 6(3):77–83, 2009.

[4] Thomas Berg, Jiongxin Liu, Seung Woo Lee, et al. Birdsnap: Large-scale fine-grained visual categorization of birds. In *CVPR*, pages 2011–2018, 2014.

[5] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. MVTec AD–A comprehensive real-world dataset for unsupervised anomaly detection. In *CVPR*, pages 9592–9600, 2019.

[6] David Berthelot, Nicholas Carlini, Ian Goodfellow, et al. MixMatch: A holistic approach to semi-supervised learning. *NeurIPS*, 32, 2019.

[7] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *ICML*, pages 1597–1607. PMLR, 2020.

[8] Yue Chen, Yalong Bai, Wei Zhang, and Tao Mei. Destruction and construction learning for fine-grained image recognition. In *CVPR*, pages 5157–5166, 2019.

[9] Marius Cordts, Mohamed Omran, Sebastian Ramos, et al. The cityscapes dataset for semantic urban scene understanding. In *CVPR*, pages 3213–3223, 2016.

[10] Jesse Davis and Mark Goadrich. The relationship between Precision-Recall and ROC curves. In *ICML*, pages 233–240, 2006.

[11] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (VOC) challenge. *IJCV*, 88(2):303–338, 2010.

[12] Christoph Feichtenhofer, Haoqi Fan, Jitendra Malik, and Kaiming He. SlowFast networks for video recognition. In *ICCV*, pages 6202–6211, 2019.

[13] Chuanxing Geng, Sheng-jun Huang, and Songcan Chen. Recent advances in open set recognition: A survey. *TPAMI*, 2020.

[14] Iman Golpour, RA Chayjan, et al. Identification and classification of bulk paddy, brown, and white rice cultivars with colour features extraction using image analysis and neural network. *Czech Journal of Food Sciences*, 32(3):280–287, 2014.

[15] Jose D Guzman, Engelbert K Peralta, et al. Classification of philippine rice grains using machine vision and artificial neural networks. In *World conference on Agricultural information and IT*, volume 6, pages 41–48, 2008.

[16] Joakim Bruslund Haurum and Thomas B Moeslund. Sewer-ML: A multi-label sewer defect classification dataset and benchmark. In *CVPR*, pages 13456–13467, 2021.

[17] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *CVPR*, pages 9729–9738, 2020.

[18] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, pages 770–778, 2016.

[19] Dan Hendrycks, Mantas Mazeika, Saurav Kadavath, and Dawn Song. Using self-supervised learning can improve model robustness and uncertainty. *NeurIPS*, 32:15663–15674, 2019.

[20] ISO 24333: Cereals and cereal products — Sampling. Standard, International Organization for Standardization, Dec. 2009.

[21] ISO 5527: Cereals – Vocabulary. Standard, International Organization for Standardization, Feb. 2015.

[22] Ying Jin, Ximei Wang, Mingsheng Long, and Jianmin Wang. Minimum class confusion for versatile domain adaptation. In *ECCV*, pages 464–480, 2020.

[23] Glenn Jocher, Alex Stoken, Ayush Chaurasia, et al. ultralytics/yolov5: v6.0 - YOLOv5n 'Nano' models, Roboflow integration, TensorFlow export, OpenCV DNN support, 2021.

[24] Andreas Kamilaris and Francesc X Prenafeta-Boldú. Deep learning in agriculture: A survey. *Computers and electronics in agriculture*, 147:70–90, 2018.

[25] Will Kay, Joao Carreira, Karen Simonyan, et al. The kinetics human action video dataset. *arXiv preprint arXiv:1705.06950*, 2017.

[26] Tsung-Yi Lin, Michael Maire, Serge Belongie, et al. Microsoft COCO: Common objects in context. In *ECCV*, pages 740–755, 2014.

[27] Ze Liu, Yutong Lin, Yue Cao, et al. Swin Transformer: Hierarchical vision transformer using shifted windows. In *ICCV*, pages 10012–10022, 2021.

[28] Ziwei Liu, Zhongqi Miao, Xiaohang Zhan, et al. Large-scale long-tailed recognition in an open world. In *CVPR*, pages 2537–2546, 2019.

[29] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *CVPR*, pages 3431–3440, 2015.

[30] Mingsheng Long, Zhangjie Cao, Jianmin Wang, and Michael I Jordan. Conditional adversarial domain adaptation. In *NIPS*, pages 1647–1657, 2018.

[31] Weiqing Min, Zhiling Wang, Yuxin Liu, et al. Large scale visual food recognition. *CoRR*, abs/2103.16107, 2021.

[32] Adam Paszke, Sam Gross, Francisco Massa, et al. Pytorch: An imperative style, high-performance deep learning library. *NeurIPS*, 32:8026–8037, 2019.

[33] Tom Pearson. Hardware-based image processing for high-speed inspection of grains. *Computers and electronics in agriculture*, 69(1):12–18, 2009.

[34] Zhengjun Qiu, Jian Chen, Yiying Zhao, et al. Variety identification of single rice seed using hyperspectral imaging combined with convolutional neural network. *Applied Sciences*, 8(2):212, 2018.

[35] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. In *NIPS*, volume 28, pages 91–99, 2015.

[36] Lukas Ruff, Robert Vandermeulen, Nico Goernitz, et al. Deep one-class classification. In *ICML*, pages 4393–4402, 2018.

[37] Olga Russakovsky, Jia Deng, Hao Su, et al. Imagenet large scale visual recognition challenge. *IJCV*, 115(3):211–252, 2015.

[38] Kuniaki Saito, Kohei Watanabe, Yoshitaka Ushiku, and Tatsuya Harada. Maximum classifier discrepancy for unsupervised domain adaptation. In *CVPR*, pages 3723–3732, 2018.

[39] Sanjivani Shantaiya and Uzma Ansari. Identification of food grains and its quality using pattern classification. In *IEEE International Conference on Communication Technology, Raipur, India*, 2010.

[40] Pei Sun, Henrik Kretzschmar, Xerxes Dotiwalla, et al. Scalability in perception for autonomous driving: Waymo open dataset. In *CVPR*, pages 2446–2454, 2020.

[41] Jihoon Tack, Sangwoo Mo, Jongheon Jeong, and Jinwoo Shin. CSI: Novelty detection via contrastive learning on distributionally shifted instances. *NeurIPS*, 33:11839–11852, 2020.

[42] Nima Tajbakhsh, Laura Jeyaseelan, Qian Li, et al. Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. *Medical Image Analysis*, 63:101693, 2020.

[43] Quin Thames, Arjun Karpur, Wade Norris, et al. Nutrition5k: Towards automatic nutritional understanding of generic food. In *CVPR*, pages 8903–8911, 2021.

[44] Neeraj Singh Visen, Jitendra Paliwal, Digvir Jayas, and NDG White. Image analysis of bulk grain samples using neural networks. In *2003 ASAE Annual Meeting*, page 1. American Society of Agricultural and Biological Engineers, 2003.

[45] P Vithu and JA Moses. Machine vision system for food grain quality evaluation: A review. *Trends in Food Science & Technology*, 56:13–20, 2016.

[46] YN Wan, CM Lin, JF Chiou, et al. Adaptive classification method for an automatic grain quality inspection system using machine vision and neural network. *ASAE Annual International Meeting*, pages 1–19, 2000.

[47] Linjie Yang, Ping Luo, Chen Change Loy, and Xiaoou Tang. A large-scale car dataset for fine-grained categorization and verification. In *CVPR*, pages 3973–3981, 2015.

[48] Piotr Zapotoczny. Discrimination of wheat grain varieties using image analysis and neural networks. part i. single kernel texture. *Journal of Cereal Science*, 54(1):60–68, 2011.

[49] Bolei Zhou, Aditya Khosla, Agata Lapedriza, et al. Learning deep features for discriminative localization. In *CVPR*, pages 2921–2929, 2016.