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

Supplementary material

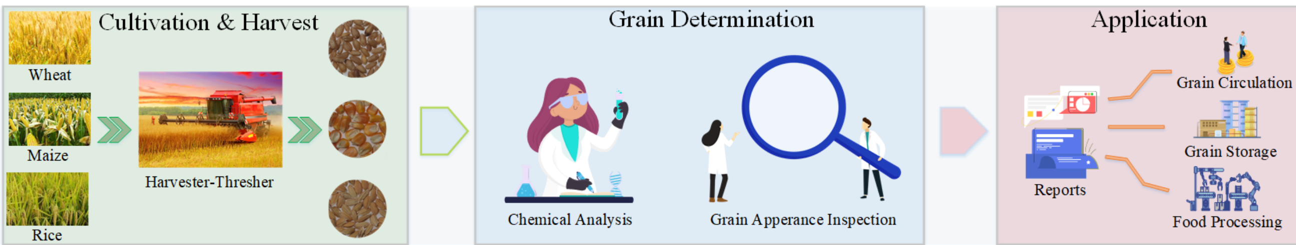

Figure I: The role of Grain Appearance Inspection (GAI).

In this supplementary material, we describe *GrainSpace*, our benchmark and grain image examples in further detail. This document includes:

- Details and statistical information about *GrainSpace* (Section A).
- Detailed results of our benchmark (Section B).
- Examples of single-kernel images of wheat, maize and rice grains (Section C).

A. *GrainSpace*

In this section, we describe more details about GAI (see Figure I) including device prototypes, data processing procedure and the statistical information of *GrainSpace*.

Figure II gives the information about the original locations of cultivation of all grain samples. We tried our best to construct *GrainSpace* with comprehensive grain samples in terms of species and regions. To avoid potential ethical issues or privacy restrictions, information about specified regions is erased. Nonetheless, we found that all grain samples are acquired from 5 countries including The United States (USA), Canada (CAN), Australia (AU), Cambodia (KHM) and China (CHN). Specifically, wheat grains are mainly sampled from USA, CAN, AU and CN; maize grains are mainly sampled from USA and CN; and rice grains are mainly sampled from KHM and CN.

Figure III shows more information about three kinds of device prototypes in terms of design blueprint, real images, raw image I_{raw}, single-kernel image I_g and radar map. At the beginning of designing device prototypes, we considered that how to capture the whole outside appearance information of grain kernels with the lowest cost but the highest efficiency. However, there is no such thing as a free

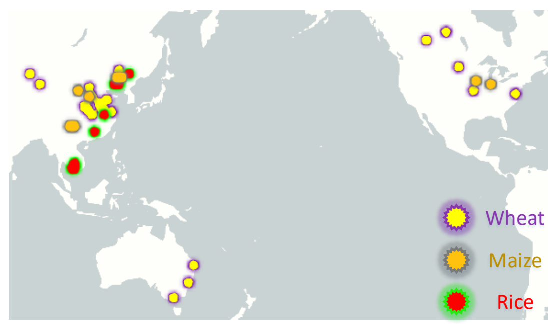

Figure II: The cultivation locations of all grain samples.

lunch. We designed P600 with an automatic feeding system and stable photograph environments, and also designed M600 with a smartphone and a simple holder. P600 with high manufacturing cost is suitable for laboratory or industrial situations, whereas M600 with low efficiency has more potential for widespread acceptance. In addition, we attempted to design G600 as an intermediate device which is a trade-off between P600 and G600. We believe that data captured by G600 can provide abundant information for training M600 models.

Figure IV describes detailed overview of data processing procedure. Taking wheat sample as an example, raw wheat grain samples obtained from granaries or freighters contain various impurities including extraneous and inorganic matters, and thus raw samples are firstly cleaned manually by using combinations of sieves with different apertures. Then, wheat samples are still mixed with impurities that have similar shape and weight with wheat kernels, in

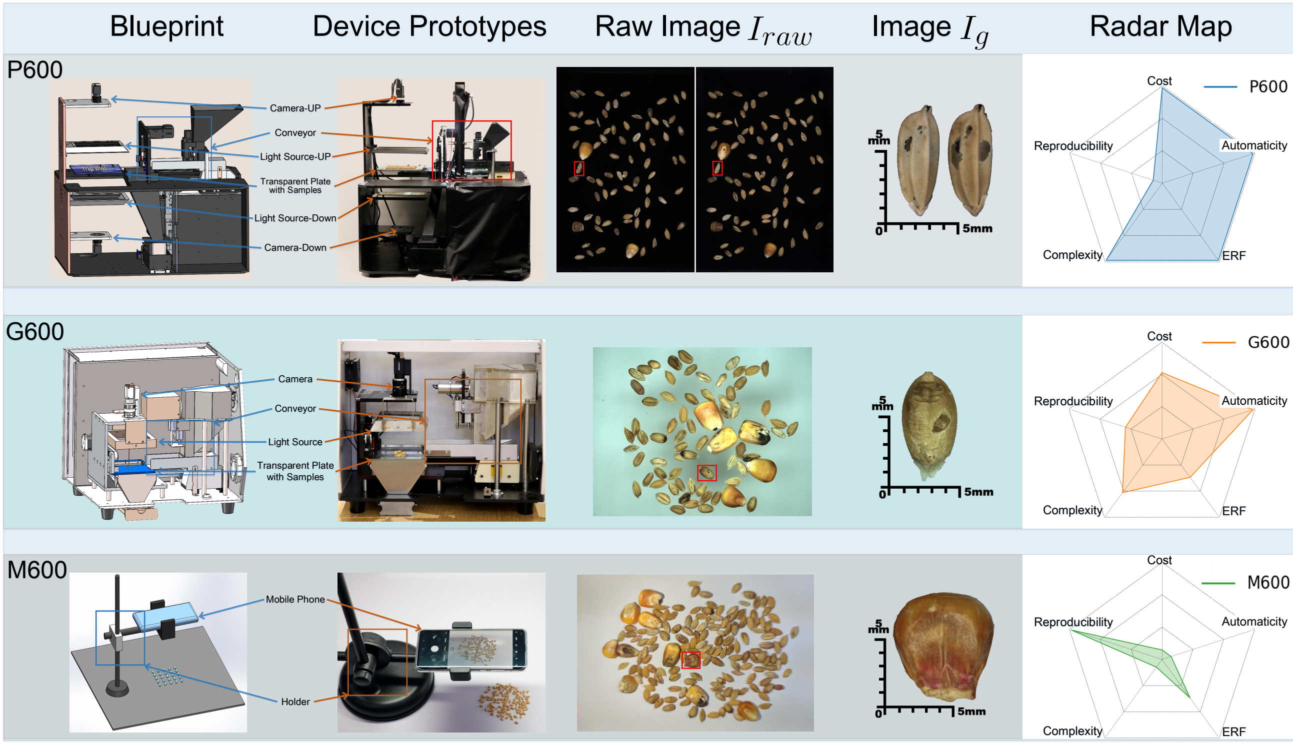

Figure III: Detailed description of three device prototypes (P600, G600 and M600).

which these impurities are picked out by inspectors with the aid of tweezers. With the above operations, test samples are obtained and then divided into several predefined groups with the inspectors' determination, each of which groups is delivered into devices in batch to obtain N raw images $I^1_{raw}, \dots, I^N_{raw}$. After capturing all raw images, raw image I_{raw} is processed to generate many single-kernel images I_g via detection and alignment stages, in which detection stage is introduced to localize and crop single-kernel images and alignment stage is used to normalize the view of single-kernel image. Finally, all aligned singe-kernel images are inspected manually to construct high-quality data.

Figure V illustrates the distribution of NORMAL and UD categories of wheat and maize data, and all sub-types of rice data. We observed that, except for NORMAL category, SD and MY are the most common UD-grains in wheat and maize respectively, whereas F&S and SD are the rarest UD-grains. These UD categories distribution are consistent with real-world situation where these kinds of UD samples are abundant or unusual.

B. Detailed Results of Benchmark

In this section, we give some backgrounds about three challenges, and describe detailed experimental settings, results and visualization of our benchmark.

Fine-grained Recognition: Fine-grained visual classification (FGVC) aims at find subtle differences among a set of classes, in which these differences are usually defined by experts. There are several common datasets, *e.g.* bird categories [40, 35], dog species, flower types [1, 25, 29], aircraft [24] and etc [19]. Recently, FGVC has attracted attention and great progress has been obtained [8, 38, 41, 43, 7, 9, 17, 20]. Due to the limitation of computational resources, in this paper, we employed DCL [8] as one of the benchmarking models. Compared to FGVC models, we also employ two common models: ResNet50 (R50) [14] and Swin Transformer [21].

- R50 models were trained with cross entropy loss for 40 epochs, batch size is 64, and we employed Radam optimizer with ReduceLROnPlateau learning rate scheduler and initial learning rate of 0.0003. All input images are resized to 224×224 and a combination of data augmentation (RandomBright, RandomContrast, RandomFlip) is adopted.
- Following original paper, DCL models were trained for 40 epochs, batch size is 48, and we employed SGD optimizer with step learning rate scheduler (halved after 10, 20 and 30 epochs) and initial learning rate of

0.001. All input images are resized to 224 × 224 and data augmentation (RandomSwap) is adopted.

- Swin Transformer models were trained with cross entropy loss for 40 epochs, batch size is 24, and we employed Radam optimizer with ReduceLROnPlateau learning rate scheduler and initial learning rate of 0.0003. All input images are resized to 224 × 224 and a combination of data augmentation (RandomBright, RandomContrast, RandomFlip) is adopted.

Table I shows the detailed results of Table 8. We used different color boxs to denote collapsed($<5\%$), limited ($<50\%$) and good ($>90\%$) performance.

In addition, we employed t-SNE [34] to qualitatively visualize features extracted from pretrained models. Figure VI presents the t-SNE visualization on all P600 data based on features extracted from DCL models. Specifically, the same classes in wheat R_{1-14} are clustered well, whereas the NORMAL and SD classes are mixed in wheat R_{15-18} and R_{19-22}. The majority of classes in maize are gathered except for a slight confusion between MY and HD. The interclasses in rice are separated clearly.

Semi-supervised Learning: We employed MixMatch [4] to fine-tune corresponding models. Models were finetuned for 40 epochs, batch size is 128, and we employed Adam optimizer with the step learning rate scheduler (halved after 10, 20 and 30 epochs) and initial learning rate of 0.00005. All input images are resized to 224 × 224 and data augmentation (Mixup [39]) is adopted. Table II shows the detailed results of Table 9.

Self-supervised Learning: We employed MoCo [13] to train models without annotations. Following the original paper, models were trained for 40 epochs (due to the limitation of GPU resources), batch size is 256, MoCo momentum is 0.999, and we employed SGD optimizer with the step learning rate scheduler (halved after 30 epochs) and initial learning rate of 0.02. All input images are resized to 224 × 224 and a combination of data augmentation (RandomResizeCrop, ColorJitter, RandomGrayscale, GaussianBlur and RandomHorizontalFlip) is adopted.

Then, we evaluated the pretrained models via linear probe with different proportions of labeled data. We extracted frozen features from pretrained models and added a linear classifier. Models are fine-tuned for 50 epochs, batch size is 256, and we employed SGD optimizer with the step learning rate scheduler (halved after 30 epochs) and initial learning rate of 5. All input images are resized to 224 × 224 and a combination of data augmentation (RandomResizeCrop and RandomHorizontalFlip) is adopted. Figure VII and Table III illustrate the detailed results of Table 10.

Domain Adaptation: The task of Domain Adaptation (DA) aims at learning models that can eliminate the distribution shift between training and testing datasets. There are serveral common datasets [37, 26] and advanced methods [5, 11, 28, 36]. In this paper, we employed CDAN [22], MCD [32] and MCC [18] to evaluate DA performance.

- CDAN models were trained for 30 epochs, batch size is 48, and we employed SGD optimizer with the specified rate scheduler and initial learning rate of 0.001. All input images are resized to 224 × 224 and a combination of data augmentation (RandomResizedCrop, RandomHorizontalFlip) is adopted.
- MCD models were trained for 30 epochs, batch size is 36, and we employed Adam optimizer with a constant learning rate of 0.0002. All input images are resized to 224 × 224.
- CDAN models were trained for 30 epochs, batch size is 48, and we employed SGD optimizer with the specified rate scheduler and initial learning rate of 0.001. All input images are resized to 224 × 224 and a combination of data augmentation (RandomResizedCrop, RandomHorizontalFlip) is adopted.

Table IV shows the detailed results of Table 11. Table V, Table VI and Table VII show the detailed results of Table 12. We observed that models adapting between G600 and M600 on wheat and rice data achieved comparable results, which verified that data acquired by G600 benefits M600 models.

Out-of-distribution Recognition: The task of Out-of-distribution (OOD) recognition (also named as an anomaly detection) aims at identifying weather a sample come from training dataset or not. There are several datasets [6, 27, 10, 3] and many deep learning-based methods [30, 23, 2, 12, 15]. In this paper, we employed Deep SVDD [31], Rot [16] and CSI [33] to evaluate out-of-distribution (OOD) recognition performance.

- Deep SVDD models were trained for 40 epochs, batch size is 128, and we employed Adam optimizer with step rate scheduler (halved after 10, 20 and 30 epochs) and initial learning rate of 0.0001. All input images are resized to 96 × 96 and data augmentation (global contrast normalization) is adopted.
- Rot models were trained for 40 epochs, batch size is 64, and we employed SGD optimizer with a constant learning rate of 0.1. All input images are resized to 96 × 96 and data augmentation (RandomResizedCrop) is adopted.
- CSI models were trained for 40 epochs, batch size is 24, and we employed SGD optimizer with the cosine rate scheduler and initial learning rate of 0.1. All input images are resized to 224 × 224 and a combination of data augmentation (RandomResizedCrop RandomHorizontalFlip) is adopted.

Table VIII shows OOD method results on G600 and M600 rice data, and Table IX shows OOD method results on G600 and M600 wheat and maize data. We observed that Rot models on wheat (M600) and maize(G600) data achieved AUROC of near 80%, which verified that treating UD-grains recognition as OOD recognition is feasible and potential.

C. Examples of Wheat, Maize and Rice Grains

We further plot single-kernel images of all categories of wheat, maize and rice grains (see Figure VIII, Figure IX and Figure X).

Figure XI shows the raw images acquired by P600, G600 and M600. We show more CAM-based [42] visualization results (see Figure XII, Figure XIII and Figure XIV).

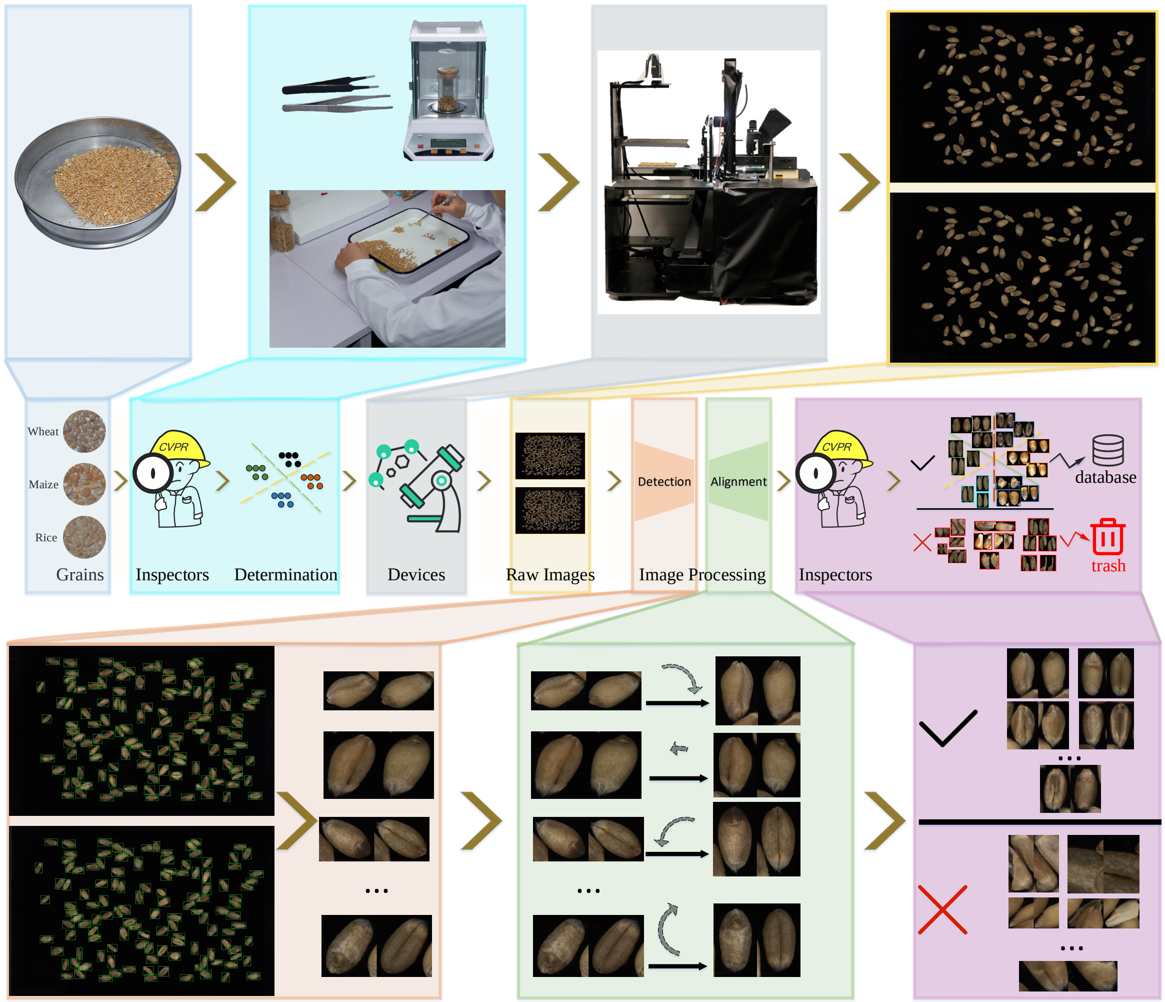

Figure IV: Detailed overview of data acquisition.

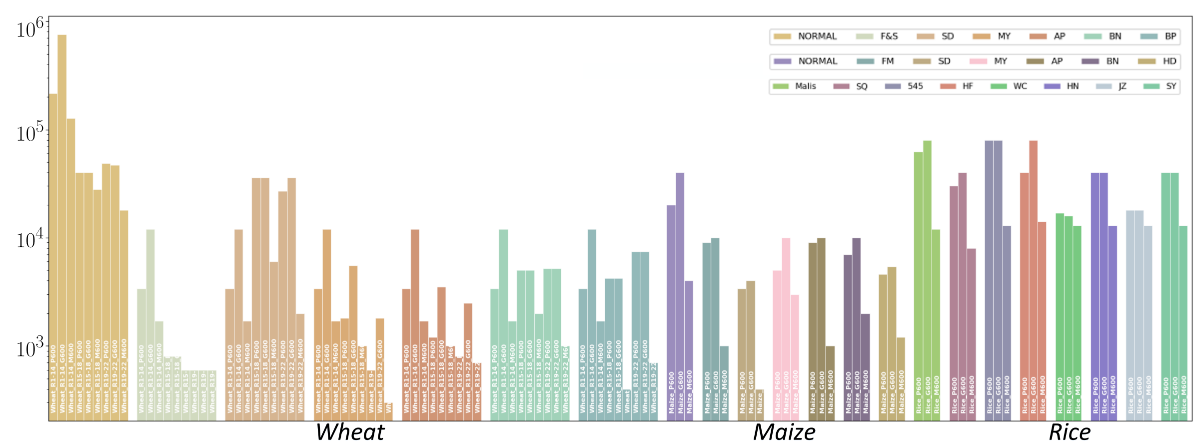

Figure V: The histogram of three kinds of grain data (wheat, maize and rice) and their subclasses.

Table I: Detailed performance of R50, DCL and Swin Transformer on wheat data: regions vs. device prototypes (supplements Table 8, all results are F1-scores, <5%, <50%, >90%).

Model	Region	Device	NORMAL	F&S	SD	MY	BN	AP	BP	Total
R50	R_{1-14}	P600	99.5%	85.2%	93.0%	93.6%	98.5%	95.5%	92.0%	**93.9%**
		G600	99.0%	95.1%	69.7%	77.0%	95.2%	70.2%	54.7%	**80.1%**
		M600	99.5%	87.1%	86.1%	80.1%	94.7%	78.5%	87.2%	87.6%
	R_{15-18}	P600	82.3%	35.9%	80.0%	93.9%	98.1%	83.4%	86.4%	80.0%
		G600	75.7%	25.9%	73.5%	96.4%	97.7%	89.8%	76.7%	76.5%
		M600	96.8%	41.4%	95.8%	94.5%	92.8%	47.4%	89.2%	**79.7%**
	R_{19-22}	P600	86.7%	34.1%	78.2%	40.0%	96.2%	76.8%	78.9%	70.1%
		G600	84.0%	38.4%	81.6%	65.0%	96.5%	94.3%	72.7%	**76.1%**
		M600	96.7%	56.6%	82.8%	60.4%	85.9%	62.8%	87.4%	**76.1%**
DCL	R_{1-14}	P600	99.4%	83.0%	92.1%	93.0%	97.7%	92.0%	90.6%	92.5%
		G600	99.0%	93.1%	69.7%	78.1%	93.9%	69.4%	50.5%	79.1%
		M600	99.5%	85.3%	85.4%	80.0%	96.1%	79.2%	89.6%	**87.9%**
	R_{15-18}	P600	82.1%	49.3%	80.3%	93.5%	98.3%	84.2%	86.9%	**82.1%**
		G600	75.1%	30.1%	72.7%	96.5%	97.1%	91.0%	77.6%	**77.2%**
		M600	96.2%	53.4%	93.6%	93.8%	87.7%	25.7%	82.5%	76.1%
	R_{19-22}	P600	86.4%	46.8%	77.5%	42.9%	97.1%	85.9%	80.6%	**73.9%**
		G600	83.1%	37.4%	79.9%	61.0%	96.5%	94.1%	72.0%	74.9%
		M600	96.0%	44.0%	78.3%	52.0%	86.7%	61.0%	88.7%	72.4%
SwinT	R_{1-14}	P600	97.4%	43.2%	20.7%	81.2%	87.5%	25.9%	40.0%	56.5%
		G600	97.0%	41.6%	17.3%	22.6%	71.2%	0.3%	24.5%	39.2%
		M600	98.8%	66.3%	52.3%	33.5%	79.8%	41.7%	75.9%	64.0%
	R_{15-18}	P600	65.0%	0.0%	54.1%	75.3%	94.1%	0.0%	59.9%	49.8%
		G600	65.1%	0.0%	50.0%	85.2%	92.3%	66.2%	50.8%	58.5%
		M600	90.0%	0.0%	69.9%	51.1%	63.0%	0.0%	33.3%	43.9%
	R_{19-22}	P600	76.2%	0.0%	64.1%	0.0%	92.8%	13.2%	61.6%	44.0%
		G600	74.6%	0.0%	69.6%	31.0%	90.8%	43.9%	49.3%	51.3%
		M600	93.5%	35.1%	64.8%	20.0%	71.7%	10.4%	78.5%	53.4%

Table II: Detailed performance of device prototypes on wheat, maize and rice data. (+ and - denote results obtained from MixMatch, supplements Table 9, all results are F1-scores, <5%, <50%, >90%).

Species	Training set P600	Training set G600	Training set M600	Test set	NORMAL	Damaged and Unsound grains: F&S	SD	MY	BN	AP	BP	Total
Wheat	✓			P600	89.4%+2.8%	67.2%+0.5%	19.8%+41.9%	67.1%+15.6%	94.9%+0.4%	73.8%+2.4%	67.6%+10.9%	68.5%+10.7%
		✓		G600	93.7%+0.8%	89.9%-4.5%	17.0%+28.3%	59.7%+4.6%	86.5%-0.7%	54.2%-0.1%	43.6%+7.3%	63.5%+5.2%
			✓	M600	95.1%+1.7%	62.3%-3.4%	32.9%+36.5%	54.3%+10.9%	57.6%+17.7%	48.7%-2.5%	65.1%+14.1%	59.4%+10.7%
		✓	✓	G600	93.7%+0.6%	89.3%-7.4%	19.3%+21.3%	62.1%+3.9%	81.8%+5.0%	56.8%-3.0%	41.1%+10.9%	63.4%+4.5%
		✓	✓	M600	88.1%+5.3%	0.1%+27.1%	0.2%+7.2%	0.1%+22.1%	10.4%+5.1%	0.5%+11.0%	4.5%+25.3%	14.8%+14.7%

Species	Training set P600	Training set G600	Training set M600	Test set	NORMAL	Damaged and Unsound grains: FM	SD	MY	BN	AP	HD	Total
Maize	✓			P600	99.2%-1.5%	99.4%-2.1%	93.6%-5.7%	86.5%-2.8%	98.0%-1.5%	96.8%-1.0%	84.8%-4.1%	94.0%-2.6%
		✓		G600	97.9%-1.4%	85.5%-2.4%	88.9%-1.5%	75.3%-6.0%	95.4%-0.6%	92.2%-3.1%	70.8%-0.3%	86.6%-2.2%
			✓	M600	95.4%-1.7%	52.4%-8.3%	85.3%-24.3%	81.3%-4.1%	94.9%-1.3%	91.4%-0.3%	79.0%-4.6%	82.8%-6.4%
		✓	✓	G600	97.7%-1.0%	83.2%-1.6%	87.5%-2.0%	72.7%-3.8%	95.7%-1.4%	91.7%-1.9%	68.4%+0.6%	85.3%-1.6%
		✓	✓	M600	67.8%+8.9%	14.4%+13.1%	27.3%+28.3%	31.4%+24.2%	62.2%+19.3%	31.7%+33.9%	1.6%+42.8%	33.8%+24.3%

Species	Training set P600	Training set G600	Training set M600	Test set	Categories of Rice Grains: Malis	SQ	545	HF	WC	HN	JZ	SY	Total
Rice	✓			P600	99.2%	97.4%	98.9%	99.0%	99.9%	99.4%	99.7%	99.8%	99.2%
		✓		G600	99.4%	95.6%	97.8%	99.4%	99.7%	99.5%	99.6%	99.8%	98.9%
			✓	M600	93.4%	80.0%	88.2%	91.6%	98.0%	97.1%	98.6%	96.8%	93.0%
		✓	✓	G600	99.3%	95.0%	97.5%	99.4%	99.7%	99.4%	99.6%	99.8%	98.7%
		✓	✓	M600	4.3%	30.0%	19.8%	21.7%	43.5%	23.0%	46.6%	25.7%	26.8%

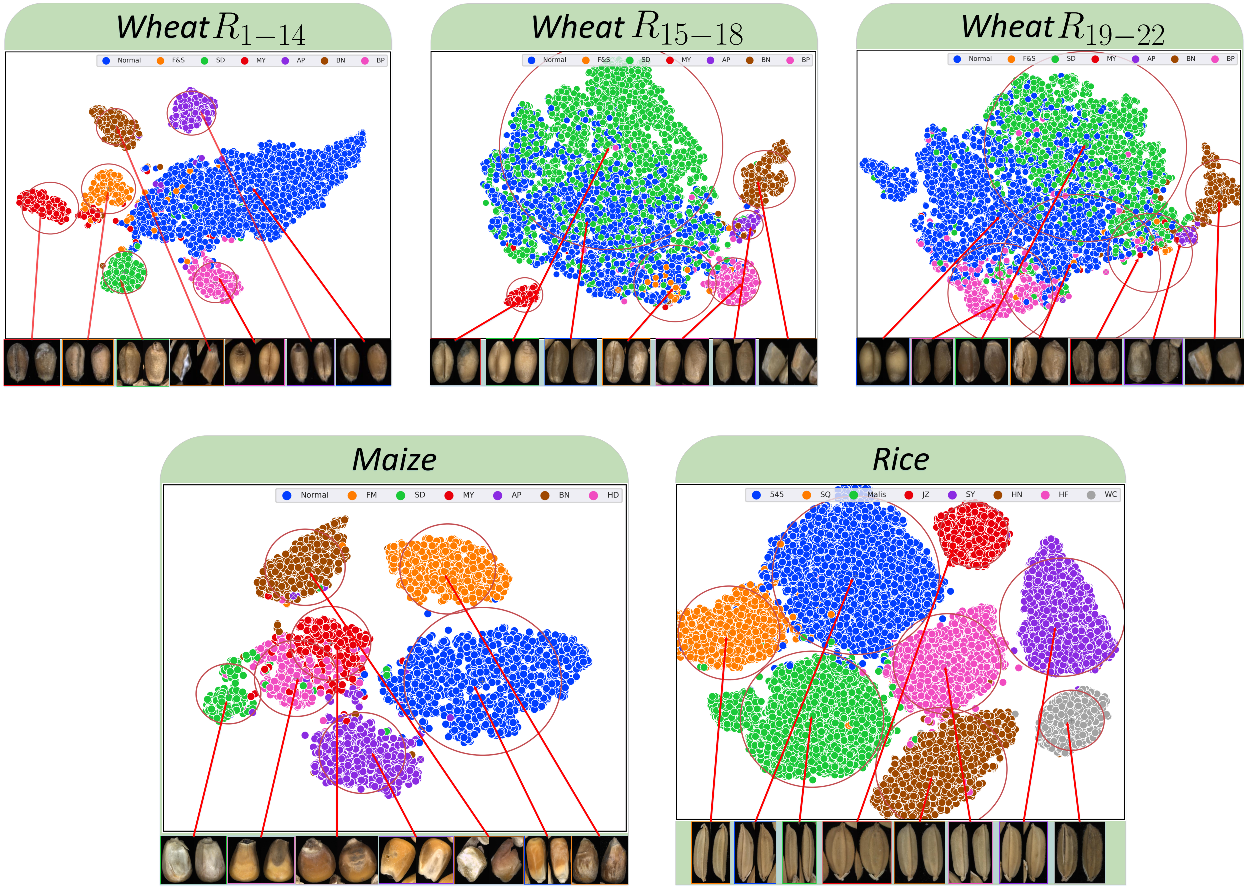

Figure VI: t-SNE visualization of P600 wheat, maize and rice data (based on DCL models).

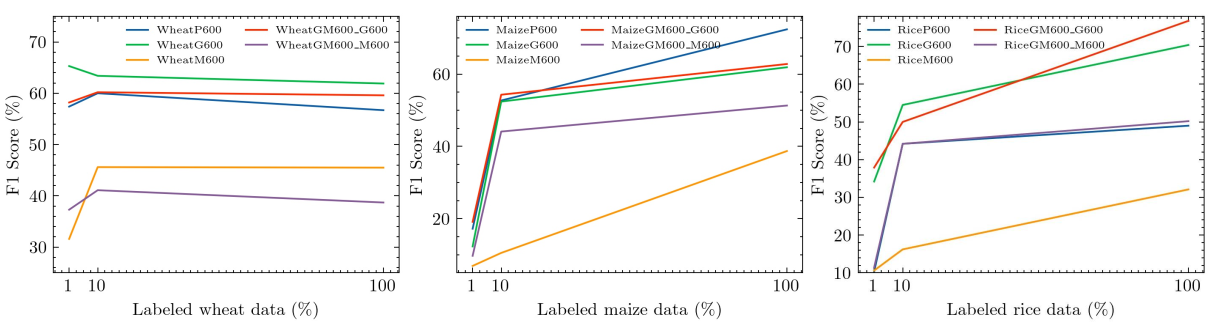

Figure VII: Curves of MoCo performance via linear probe vs. different proportions of labeled data.

Table III: Detailed MoCo performance of device prototypes on wheat, maize and rice data (supplements Table 10, all results are F1-scores, <5%, <50%, >90%).

Species	Training set			Test set	Proportion	NORMAL	Damaged and Unsound Grains						Total
	P600	G600	M600				F&S	SD	MY	BN	AP	BP	
Wheat	✓			P600	1%	89.7%	19.0%	60.3%	69.5%	96.1%	27.6%	39.8%	57.4%
	✓				10%	90.2%	35.2%	59.8%	76.3%	96.3%	22.9%	39.2%	60.0%
	✓				100%	89.8%	27.4%	55.6%	72.1%	96.1%	21.2%	34.6%	56.7%
		✓		G600	1%	94.1%	59.2%	50.4%	75.4%	94.8%	55.6%	27.8%	65.3%
		✓			10%	93.8%	51.2%	44.5%	75.4%	95.5%	57.0%	26.2%	63.4%
		✓			100%	93.7%	53.5%	40.7%	72.2%	95.4%	49.1%	28.4%	61.9%
			✓	M600	1%	94.7%	28.0%	0.0%	20.4%	70.1%	0.6%	7.2%	31.6%
			✓		10%	95.3%	52.0%	29.7%	21.2%	83.2%	25.9%	12.2%	45.6%
			✓		100%	95.5%	51.0%	37.2%	32.6%	82.7%	12.0%	7.4%	45.5%
		✓	✓	G600 M600	1%	93.2% 94.5%	52.5% 32.3%	30.7% 10.2%	73.1% 33.7%	94.7% 73.8%	37.7% 9.8%	25.7% 6.8%	58.2% 37.3%
		✓	✓		10%	93.4% 94.7%	47.8% 38.9%	34.8% 15.2%	74.4% 36.0%	95.2% 77.0%	48.8% 16.1%	26.9% 9.5%	60.2% 41.1%
		✓	✓		100%	93.3% 94.5%	49.1% 36.3%	31.7% 9.2%	73.0% 31.5%	95.1% 76.5%	48.1% 11.9%	26.9% 10.8%	59.6% 38.7%

Species	Training set			Test set	Proportion	NORMAL	Damaged and Unsound Grains						Total
	P600	G600	M600				FM	SD	MY	BN	AP	HD	
Maize	✓			P600	1%	64.1%	0.1%	0.1%	0.1%	56.1%	0.1%	0.1%	17.2%
	✓				10%	83.4%	77.9%	0.1%	31.8%	88.1%	49.9%	37.8%	52.7%
	✓				100%	86.5%	92.8%	55.0%	59.4%	92.6%	61.9%	58.3%	72.4%
		✓		G600	1%	0.1%	39.3%	0.1%	0.1%	31.1%	0.1%	15.8%	12.3%
		✓			10%	80.9%	59.7%	39.2%	44.5%	76.3%	33.8%	32.6%	52.4%
		✓			100%	86.5%	64.8%	53.3%	50.9%	84.2%	54.0%	39.7%	61.9%
			✓	M600	1%	48.2%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	6.9%
			✓		10%	67.1%	0.1%	0.1%	0.1%	0.1%	6.2%	0.1%	10.5%
			✓		100%	74.5%	14.0%	29.9%	53.1%	69.2%	28.4%	1.7%	38.7%
		✓	✓	G600 M600	1%	78.2% 48.6%	0.1% 0.1%	0.1% 0.1%	29.9% 0.1%	3.0% 0.1%	0.1% 0.1%	22.3% 19.6%	19.1% 9.7%
		✓	✓		10%	79.5% 76.0%	60.6% 26.1%	40.9% 4.9%	44.6% 56.9%	79.6% 77.4%	40.7% 39.4%	34.2% 28.1%	54.3% 44.1%
		✓	✓		100%	87.0% 80.2%	65.1% 34.3%	54.8% 38.3%	51.1% 49.2%	85.3% 78.9%	53.2% 48.7%	42.8% 29.5%	62.8% 51.3%

Species	Training set			Test set	Proportion	Categories of Rice Grains								Total
	P600	G600	M600			Malis	SQ	545	HF	WC	HN	JZ	SY	
Rice	✓			P600	1%	0.0%	0.0%	46.0%	0.0%	0.0%	0.0%	36.7%	0.0%	10.3%
	✓				10%	37.2%	13.3%	47.4%	25.6%	61.3%	38.7%	65.2%	64.1%	44.2%
	✓				100%	30.0%	16.8%	49.6%	32.5%	72.8%	40.8%	79.2%	70.2%	49.0%
		✓		G600	1%	0.0%	0.0%	73.1%	56.7%	57.8%	38.4%	47.5%	0.0%	34.2%
		✓			10%	32.7%	10.9%	46.3%	71.0%	84.6%	47.8%	75.1%	67.8%	54.5%
		✓			100%	79.2%	41.6%	57.5%	83.6%	79.2%	53.7%	88.6%	79.4%	70.4%
			✓	M600	1%	0.0%	0.0%	0.0%	15.4%	28.7%	21.8%	18.9%	0.0%	10.6%
			✓		10%	15.9%	6.5%	1.0%	24.3%	25.6%	18.8%	17.2%	19.9%	16.2%
			✓		100%	33.6%	21.8%	33.6%	23.6%	41.7%	31.8%	35.5%	35.2%	32.1%
		✓	✓	G600 M600	1%	91.1% 12.5%	0.8% 20.4%	72.8% 8.5%	90.7% 13.7%	0.0% 0.1%	47.4% 35.7%	0.0% 0.1%	0.0% 0.1%	37.9% 11.4%
		✓	✓		10%	46.9% 36.0%	5.9% 22.1%	26.4% 43.9%	73.6% 26.6%	86.3% 75.1%	46.7% 45.8%	50.2% 57.7%	63.8% 45.6%	50.0% 44.2%
		✓	✓		100%	84.4% 46.3%	42.8% 10.7%	66.9% 45.3%	87.4% 32.3%	90.4% 83.1%	57.0% 45.6%	94.9% 77.6%	90.6% 61.0%	76.8% 50.2%

Table IV: Detailed performance of DA methods on wheat data: regions vs. device prototypes (supplements Table 11, all results are F1-scores, <5%, <50%, >90%).

S → T	Method	Normal	F&S	SD	MY	BN	AP	BP	Total
R_{1-14} ↓ R_{15-18}	Source only	67.5%, 60.6%, 80.2%	0.0%, 0.0%, 2.7%	14.1%, 0.0%, 20.8%	52.5%, 0.0%, 13.3%	94.5%, 70.0%, 31.7%	25.5%, 1.7%, 0.0%	46.0%, 0.0%, 10.3%	42.9%, 18.9%, 22.7%
	CDAN	64.8%, 60.4%, 83.6%	0.0%, 0.0%, 0.0%	3.9%, 0.2%, 2.2%	4.5%, 12.9%, 2.0%	72.0%, 45.2%, 2.9%	18.1%, 2.2%, 0.0%	29.0%, 0.0%, 9.5%	27.5%, 17.3%, 14.3%
	MCD	63.7%, 59.8%, 83.6%	0.0%, 0.0%, 0.0%	0.0%, 0.0%, 0.0%	0.0%, 12.4%, 0.0%	75.9%, 0.0%, 0.0%	0.0%, 0.0%, 0.0%	0.0%, 0.0%, 0.0%	20.0%, 10.3%, 11.9%
	MCC	65.9%, 61.4%, 83.8%	0.0%, 0.0%, 0.0%	7.8%, 0.0%, 0.0%	3.6%, 17.8%, 0.0%	90.7%, 65.5%, 11.0%	18.2%, 0.6%, 0.0%	35.9%, 0.5%, 13.6%	31.8%, 20.8%, 15.5%
R_{15-18} ↓ R_{1-14}	Source only	49.7%, 5.8%, 58.2%	0.8%, 0.0%, 2.1%	8.6%, 16.6%, 7.1%	6.4%, 3.6%, 15.9%	8.6%, 72.3%, 66.7%	16.3%, 11.4%, 2.1%	33.1%, 3.6%, 6.1%	17.6%, 16.2%, 22.6%
	CDAN	87.8%, 44.3%, 85.3%	2.6%, 0.0%, 5.5%	16.2%, 1.7%, 5.7%	18.1%, 7.3%, 16.4%	66.3%, 53.2%, 39.7%	8.8%, 5.3%, 0.1%	21.8%, 9.3%, 34.8%	31.6%, 17.3%, 26.8%
	MCD	69.0%, 85.0%, 92.9%	0.0%, 0.0%, 0.0%	9.3%, 5.6%, 11.7%	20.0%, 5.5%, 3.4%	67.2%, 11.1%, 18.3%	2.0%, 0.3%, 0.0%	23.3%, 38.8%, 0.0%	27.2%, 20.9%, 18.1%
	MCC	80.9%, 50.9%, 84.1%	1.6%, 0.0%, 0.0%	12.4%, 2.2%, 4.3%	14.0%, 3.9%, 3.2%	67.5%, 36.6%, 33.2%	5.1%, 5.9%, 0.0%	14.8%, 8.1%, 19.2%	28.0%, 15.4%, 20.6%
R_{15-18} ↓ R_{19-22}	Source only	77.3%, 11.3%, 87.5%	1.9%, 3.1%, 0.0%	62.1%, 7.2%, 48.8%	19.2%, 3.8%, 42.9%	88.8%, 58.3%, 63.9%	56.8%, 10.5%, 36.0%	64.5%, 18.3%, 44.3%	52.9%, 16.1%, 46.2%
	CDAN	71.8%, 30.5%, 90.8%	0.0%, 0.0%, 0.0%	44.8%, 34.3%, 39.3%	9.2%, 4.6%, 52.5%	80.4%, 55.9%, 40.0%	35.4%, 13.4%, 10.8%	64.8%, 23.4%, 64.5%	43.7%, 23.2%, 42.6%
	MCD	72.4%, 45.4%, 87.3%	0.0%, 0.0%, 0.0%	29.5%, 18.1%, 35.5%	18.6%, 7.1%, 25.9%	84.6%, 71.7%, 46.4%	4.7%, 6.5%, 0.0%	54.6%, 21.7%, 0.0%	37.7%, 24.4%, 27.9%
	MCC	67.4%, 17.1%, 88.2%	0.0%, 0.0%, 0.0%	42.9%, 32.0%, 28.9%	10.1%, 3.1%, 0.0%	78.5%, 51.6%, 38.5%	24.9%, 12.2%, 0.0%	57.3%, 20.2%, 47.0%	40.1%, 19.5%, 28.9%
R_{19-22} ↓ R_{15-18}	Source only	69.6%, 34.9%, 89.4%	0.0%, 0.0%, 11.4%	41.5%, 15.8%, 41.5%	7.5%, 21.3%, 48.6%	95.2%, 90.2%, 72.4%	39.1%, 15.2%, 18.6%	66.5%, 9.0%, 55.4%	45.6%, 26.6%, 48.2%
	CDAN	67.0%, 23.8%, 87.5%	0.0%, 0.0%, 6.8%	29.4%, 52.1%, 28.5%	1.1%, 0.0%, 43.8%	94.2%, 46.0%, 54.6%	29.5%, 7.2%, 0.0%	65.4%, 22.9%, 54.3%	40.9%, 21.7%, 39.4%
	MCD	44.5%, 36.1%, 83.7%	0.0%, 0.0%, 0.0%	51.2%, 51.1%, 22.3%	0.0%, 0.0%, 0.0%	94.2%, 76.0%, 4.3%	0.0%, 6.3%, 0.0%	65.4%, 38.9%, 22.2%	36.4%, 29.8%, 19.0%
	MCC	64.4%, 50.3%, 85.7%	0.0%, 0.0%, 5.1%	18.6%, 34.8%, 19.2%	0.0%, 0.0%, 2.0%	87.7%, 39.5%, 38.8%	3.2%, 3.1%, 0.0%	57.9%, 37.2%, 43.8%	33.1%, 23.6%, 27.8%
R_{19-22} ↓ R_{1-14}	Source only	80.3%, 60.1%, 68.7%	22.1%, 15.4%, 8.7%	7.8%, 20.1%, 5.1%	0.6%, 2.2%, 9.6%	24.7%, 92.1%, 26.8%	3.1%, 22.2%, 1.7%	44.3%, 21.6%, 28.8%	26.1%, 33.4%, 21.3%
	CDAN	83.7%, 61.9%, 93.4%	25.2%, 0.0%, 18.5%	7.7%, 0.4%, 13.3%	0.0%, 0.0%, 5.0%	66.7%, 48.7%, 51.8%	12.8%, 0.3%, 1.1%	44.9%, 55.6%, 54.8%	34.4%, 23.9%, 33.9%
	MCD	74.1%, 36.4%, 94.8%	0.0%, 0.0%, 0.0%	4.1%, 0.8%, 3.1%	0.0%, 0.0%, 0.0%	61.4%, 66.6%, 2.6%	0.0%, 0.0%, 0.0%	49.3%, 43.8%, 37.5%	27.0%, 21.1%, 19.7%
	MCC	84.7%, 57.8%, 89.6%	0.0%, 0.0%, 16.1%	1.9%, 0.4%, 6.1%	0.0%, 0.0%, 0.0%	58.2%, 39.3%, 19.1%	0.0%, 3.9%, 1.1%	34.7%, 45.6%, 15.6%	25.6%, 21.0%, 21.0%
R_{1-14} ↓ R_{19-22}	Source only	75.7%, 67.0%, 84.6%	40.4%, 11.0%, 9.1%	16.8%, 3.0%, 20.4%	3.9%, 0.0%, 6.5%	93.1%, 82.1%, 30.0%	35.4%, 5.1%, 2.3%	61.3%, 31.5%, 33.3%	46.7%, 28.5%, 26.6%
	CDAN	72.9%, 64.7%, 87.3%	14.6%, 6.3%, 3.3%	3.6%, 1.3%, 0.0%	1.1%, 9.1%, 5.7%	77.5%, 23.9%, 5.7%	30.4%, 0.8%, 0.0%	32.3%, 5.8%, 12.7%	33.2%, 16.0%, 16.4%
	MCD	71.8%, 65.0%, 86.4%	0.0%, 6.9%, 3.5%	0.1%, 0.0%, 0.0%	0.0%, 0.0%, 0.0%	75.1%, 32.0%, 0.0%	0.0%, 0.0%, 0.0%	0.0%, 0.0%, 0.0%	21.0%, 14.8%, 12.8%
	MCC	72.6%, 65.3%, 87.4%	11.4%, 2.7%, 3.2%	4.0%, 0.6%, 0.0%	0.7%, 9.4%, 0.0%	81.6%, 55.8%, 0.0%	22.6%, 0.0%, 0.0%	37.3%, 3.2%, 22.7%	32.9%, 19.6%, 16.2%

Table V: Detailed DA method performance of device prototypes on **wheat** data (supplements Table 12, all results are F1-scores, <5%, <50%, >90%).

S → T	Method	NORMAL	F&S	SD	MY	BN	AP	BP	Total
P600 ↓ G600	Source only	8.0%	25.4%	0.9%	9.2%	4.2%	21.8%	11.4%	11.6%
	CDAN	26.2%	37.6%	12.6%	9.4%	28.7%	12.4%	2.9%	18.5%
	MCD	84.8%	0.0%	6.3%	0.0%	10.1%	0.0%	0.2%	14.5%
	MCC	51.2%	1.6%	10.3%	0.0%	21.8%	0.0%	2.0%	12.4%
G600 ↓ P600	Source only	84.8%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	12.1%
	CDAN	85.0%	0.0%	0.0%	6.2%	0.1%	0.0%	0.0%	13.0%
	MCD	84.6%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	12.2%
	MCC	84.7%	0.0%	0.0%	0.0%	2.7%	0.0%	0.0%	12.5%
G600 ↓ M600	Source only	77.8%	0.0%	0.4%	0.0%	9.5%	0.6%	4.3%	13.2%
	CDAN	92.9%	0.0%	0.0%	1.3%	0.0%	0.0%	0.0%	13.4%
	MCD	92.0%	6.7%	10.2%	0.0%	0.0%	0.0%	0.0%	15.6%
	MCC	92.5%	0.0%	0.2%	0.0%	1.4%	0.0%	0.0%	13.4%
M600 ↓ G600	Source only	46.3%	8.8%	23.4%	8.9%	73.3%	9.4%	10.0%	25.7%
	CDAN	89.6%	3.4%	0.0%	0.0%	9.3%	0.0%	6.5%	15.5%
	MCD	83.3%	6.2%	5.1%	4.2%	1.1%	0.0%	13.3%	16.2%
	MCC	89.3%	1.3%	0.7%	0.4%	15.1%	0.0%	21.4%	18.3%
M600 ↓ P600	Source only	35.2%	3.4%	0.0%	0.0%	0.0%	0.0%	7.4%	6.6%
	CDAN	84.6%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	12.1%
	MCD	84.4%	0.0%	0.0%	0.0%	4.3%	0.0%	0.0%	12.7%
	MCC	84.7%	0.0%	0.0%	0.0%	0.3%	0.0%	0.0%	12.1%
P600 ↓ M600	Source only	11.8%	2.4%	1.1%	1.0%	5.6%	1.4%	7.2%	4.4%
	CDAN	82.5%	1.0%	10.7%	1.2%	9.2%	1.9%	3.1%	15.7%
	MCD	91.0%	0.0%	4.0%	0.0%	11.7%	0.0%	0.0%	15.2%
	MCC	42.3%	1.1%	13.0%	0.6%	5.7%	0.6%	3.8%	9.6%

Table VI: Detailed DA method performance of device prototypes on **maize** data (supplements Table 12, all results are F1-scores, <5%, <50%, >90%).

S → T	Method	NORMAL	FM	SD	MY	BN	AP	HD	Total
P600 ↓ G600	Source only	40.9%	23.4%	26.1%	3.1%	15.0%	25.3%	16.5%	21.5%
	CDAN	9.1%	40.2%	9.8%	26.7%	41.7%	27.3%	22.6%	25.3%
	MCD	84.1%	2.2%	1.7%	27.6%	22.6%	37.5%	15.7%	27.3%
	MCC	12.2%	32.2%	27.5%	35.3%	28.4%	27.7%	24.9%	26.9%
G600 ↓ P600	Source only	0.0%	37.4%	9.5%	0.0%	6.3%	0.2%	0.0%	7.6%
	CDAN	50.2%	53.2%	20.7%	3.7%	35.3%	43.3%	23.1%	32.8%
	MCD	76.2%	65.9%	0.0%	13.7%	55.4%	9.6%	25.0%	35.1%
	MCC	0.0%	67.3%	18.9%	21.2%	44.6%	19.7%	8.6%	25.8%
G600 ↓ M600	Source only	64.8%	8.8%	17.7%	37.2%	64.2%	16.7%	0.0%	29.9%
	CDAN	56.8%	27.7%	2.2%	13.1%	14.0%	30.9%	0.0%	20.7%
	MCD	66.1%	34.7%	0.0%	25.7%	32.6%	42.5%	9.4%	30.1%
	MCC	59.6%	28.8%	17.0%	20.3%	33.2%	24.6%	18.5%	28.9%
M600 ↓ G600	Source only	36.4%	12.4%	8.7%	31.9%	34.5%	15.5%	12.3%	21.7%
	CDAN	41.3%	39.3%	0.5%	15.6%	30.7%	0.0%	9.3%	19.5%
	MCD	66.5%	16.4%	0.0%	21.2%	48.5%	16.4%	12.0%	25.9%
	MCC	7.9%	55.0%	0.5%	32.6%	41.3%	6.2%	5.9%	21.2%
M600 ↓ P600	Source only	0.0%	38.3%	0.0%	11.5%	42.8%	0.0%	0.0%	13.2%
	CDAN	41.7%	59.2%	8.8%	25.0%	14.1%	0.0%	5.5%	22.0%
	MCD	60.6%	31.0%	5.5%	25.0%	8.1%	14.8%	9.1%	22.0%
	MCC	34.4%	53.3%	0.0%	15.3%	10.3%	0.7%	7.4%	17.3%
P600 ↓ M600	Source only	56.4%	18.4%	0.0%	0.0%	23.5%	27.2%	0.0%	17.9%
	CDAN	60.2%	5.3%	2.0%	6.7%	39.0%	11.1%	1.6%	18.0%
	MCD	58.6%	3.4%	6.2%	21.0%	12.3%	22.2%	8.7%	18.9%
	MCC	28.4%	3.2%	6.6%	3.0%	25.5%	22.8%	7.8%	13.9%

Table VII: Detailed DA method performance of device prototypes on **rice** data (supplements Table 12, all results are F1-scores, <5%, <50%, >90%).

S → T	Method	Malis	SQ	545	HF	WC	HN	JZ	SY	Total
P600 ↓ G600	Source only	10.2%	15.4%	1.3%	0.0%	0.0%	20.6%	22.3%	0.0%	8.7%
	CDAN	55.0%	11.3%	61.8%	54.8%	14.8%	2.2%	58.9%	59.2%	39.7%
	MCD	7.1%	10.4%	38.5%	32.9%	0.0%	0.0%	0.0%	19.8%	13.6%
	MCC	40.6%	1.2%	54.6%	44.4%	2.1%	1.2%	53.5%	49.1%	30.8%
G600 ↓ P600	Source only	0.2%	31.3%	39.3%	1.0%	29.1%	57.8%	35.8%	22.6%	27.1%
	CDAN	1.0%	42.7%	67.4%	7.0%	49.2%	43.9%	19.6%	52.4%	35.4%
	MCD	35.4%	0.0%	30.2%	6.1%	0.0%	14.1%	0.0%	0.0%	10.7%
	MCC	5.3%	40.2%	35.6%	11.8%	20.7%	36.5%	4.2%	34.6%	23.6%
G600 ↓ M600	Source only	2.6%	29.3%	14.0%	21.0%	26.4%	22.8%	44.2%	24.1%	23.1%
	CDAN	15.7%	21.3%	34.1%	18.5%	29.0%	21.9%	48.7%	35.4%	28.1%
	MCD	17.2%	2.9%	20.2%	13.2%	0.0%	5.4%	0.0%	15.2%	9.3%
	MCC	9.6%	26.6%	31.0%	14.8%	41.3%	37.7%	51.4%	29.1%	30.2%
M600 ↓ G600	Source only	64.1%	5.8%	71.5%	52.8%	79.1%	10.3%	90.0%	78.1%	56.5%
	CDAN	23.4%	24.7%	63.5%	11.2%	58.2%	23.7%	75.7%	51.6%	41.5%
	MCD	38.7%	0.0%	0.0%	4.7%	8.0%	5.4%	0.1%	25.6%	10.3%
	MCC	45.3%	11.5%	51.4%	14.5%	55.1%	9.1%	76.3%	56.4%	40.0%
M600 ↓ P600	Source only	0.0%	0.0%	0.2%	5.1%	7.4%	22.9%	0.1%	0.4%	4.5%
	CDAN	0.1%	14.8%	34.2%	17.5%	22.9%	20.0%	7.3%	5.3%	15.3%
	MCD	0.0%	15.4%	37.3%	24.5%	0.0%	0.0%	34.2%	0.0%	13.9%
	MCC	0.1%	14.2%	53.6%	10.7%	20.9%	28.4%	0.7%	0.5%	16.1%
P600 ↓ M600	Source only	25.2%	3.5%	4.2%	2.9%	0.0%	32.1%	21.5%	1.2%	11.3%
	CDAN	12.8%	14.3%	34.8%	20.6%	3.0%	0.9%	45.1%	29.6%	20.1%
	MCD	28.6%	10.5%	15.8%	6.5%	0.0%	0.0%	0.0%	15.4%	9.6%
	MCC	28.9%	20.6%	23.0%	16.8%	16.2%	7.8%	50.1%	27.6%	23.9%

Table VIII: OOD method performance on G600 and M600 rice data (✓ denotes this group is in-distribution).

Device	Method	Malis	SQ	545	HF	WC	HN	JZ	SY	AUROC
G600	Deep SVDD	✓	✓	✓						**67.7%**
					✓	✓	✓			47.7%
								✓	✓	57.3%
	Rot	✓	✓	✓						58.7%
					✓	✓	✓			**51.1%**
								✓	✓	56.9%
	CSI	✓	✓	✓						54.6%
					✓	✓	✓			41.3%
								✓	✓	**62.0%**
Device	Method	Malis	SQ	545	HF	WC	HN	JZ	SY	AUROC
M600	Deep SVDD	✓	✓	✓						63.3%
					✓	✓	✓			49.1%
								✓	✓	**58.3%**
	Rot	✓	✓	✓						44.1%
					✓	✓	✓			**61.3%**
								✓	✓	55.8%
	CSI	✓	✓	✓						**65.3%**
					✓	✓	✓			53.2%
								✓	✓	47.0%

Table IX: OOD method performance on G600 and M600 wheat and maize data (✓ denotes this group is in-distribution).

Species	Device	Method	Normal	F&S	SD	MY	AP	BN	BP	AUROC
Wheat	G600	Deep SVDD	✓		✓		✓	✓		52.5%
				✓		✓			✓	52.5%
		Rot	✓		✓		✓	✓		**70.0%**
				✓		✓			✓	**61.1%**
		CSI	✓		✓		✓	✓		57.8%
				✓		✓			✓	42.3%
	M600	Deep SVDD	✓		✓		✓	✓		64.0%
				✓		✓			✓	**69.3%**
		Rot	✓		✓		✓	✓		**83.1%**
				✓		✓			✓	43.5%
		CSI	✓		✓		✓	✓		72.5%
				✓		✓			✓	60.1%
Species	Device	Method	Normal	FM	SD	MY	AP	BN	HD	AUROC
Maize	G600	Deep SVDD	✓		✓		✓	✓		70.4%
				✓		✓			✓	**51.8%**
		Rot	✓		✓		✓	✓		**79.6%**
				✓		✓			✓	49.1%
		CSI	✓		✓		✓	✓		64.6%
				✓		✓			✓	44.5%
	M600	Deep SVDD	✓		✓		✓	✓		62.2%
				✓		✓			✓	**50.9%**
		Rot	✓		✓		✓	✓		**66.6%**
				✓		✓			✓	38.5%
		CSI	✓		✓		✓	✓		62.4%
				✓		✓			✓	43.9%

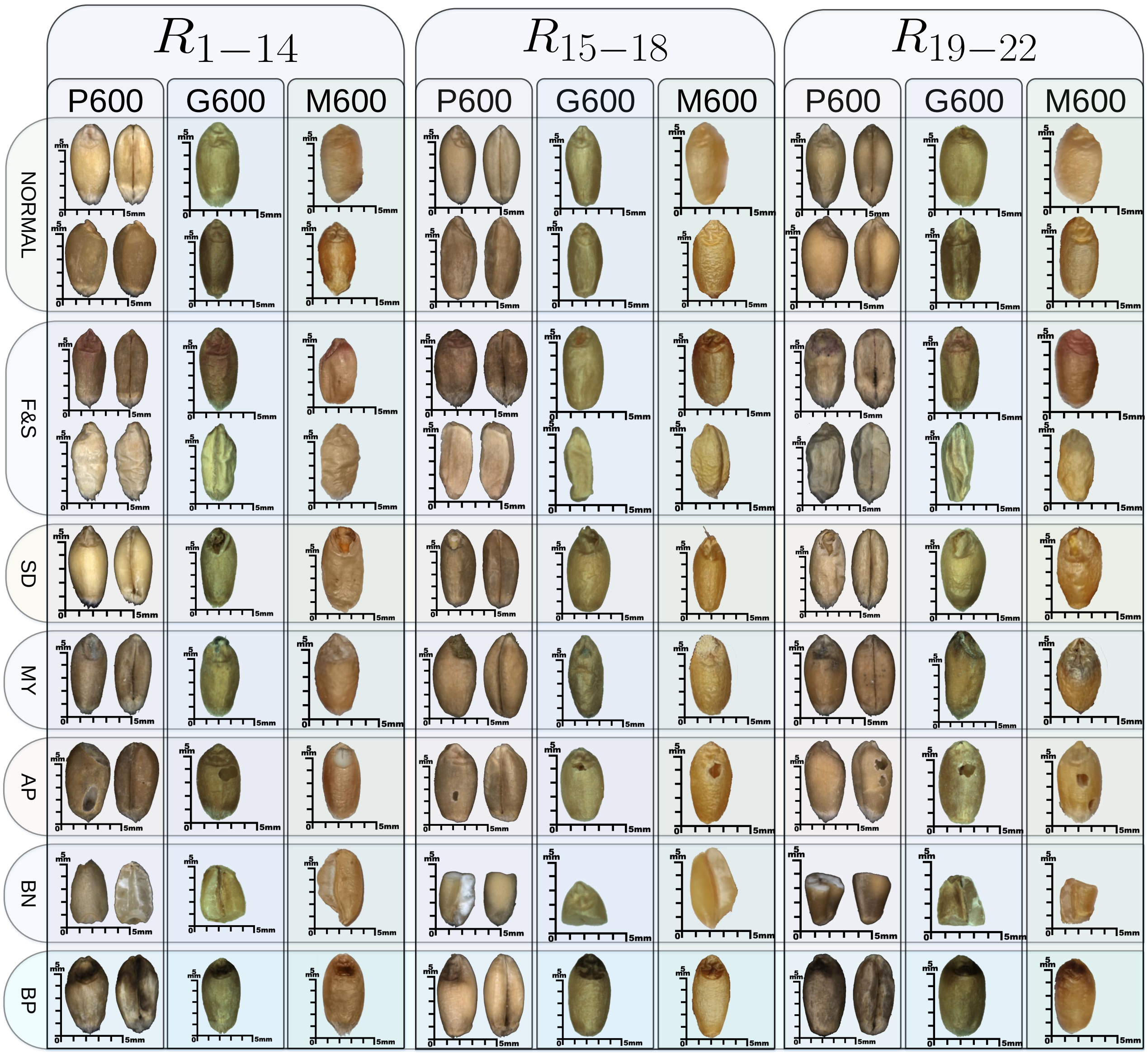

Figure VIII: Examples of single-kernel images of wheat grains.

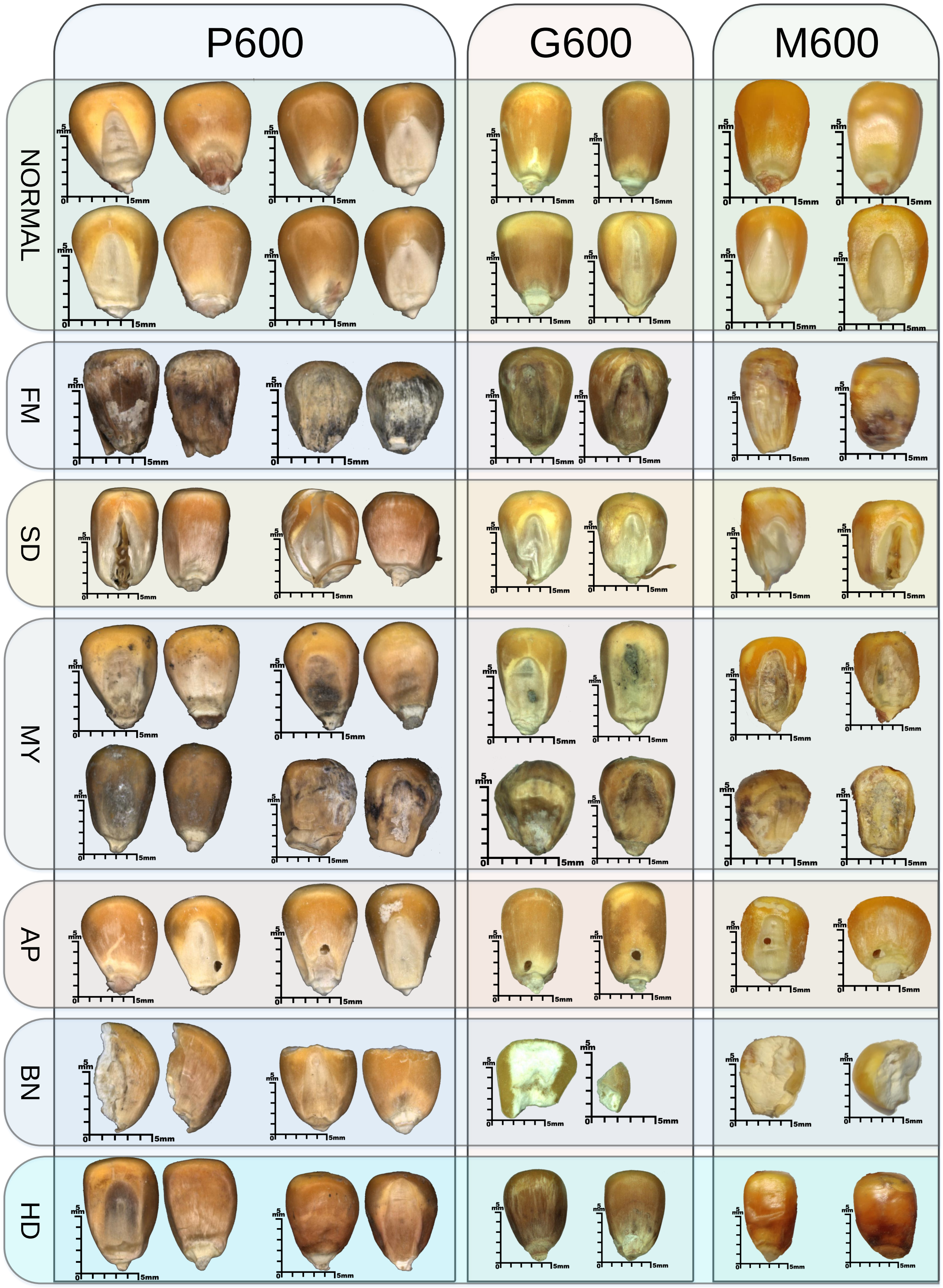

Figure IX: Examples of single-kernel images of maize grains.

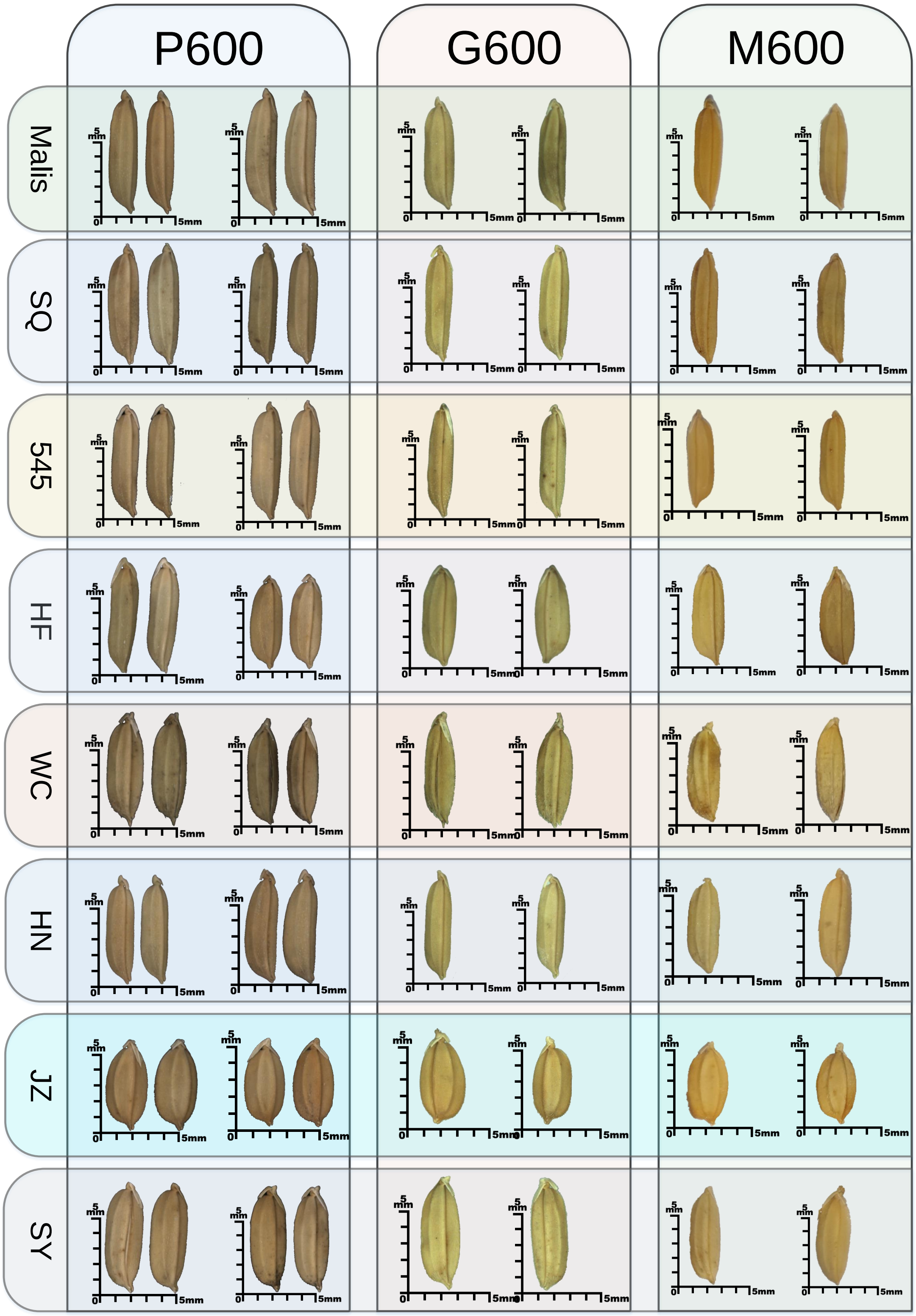

Figure X: Examples of single-kernel images of rice grains.

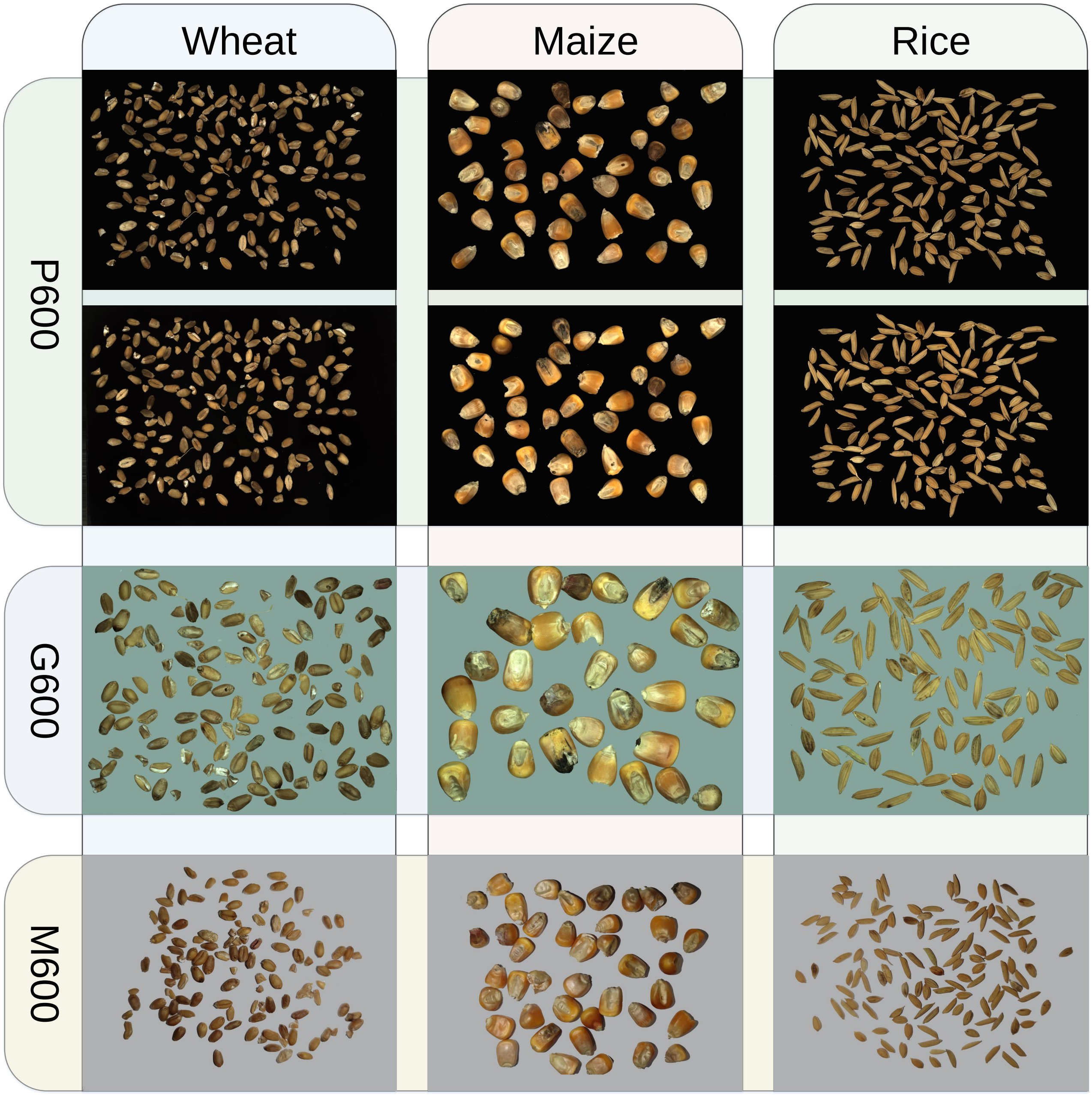

Figure XI: Examples of raw images captured by P600, G600 and M600.

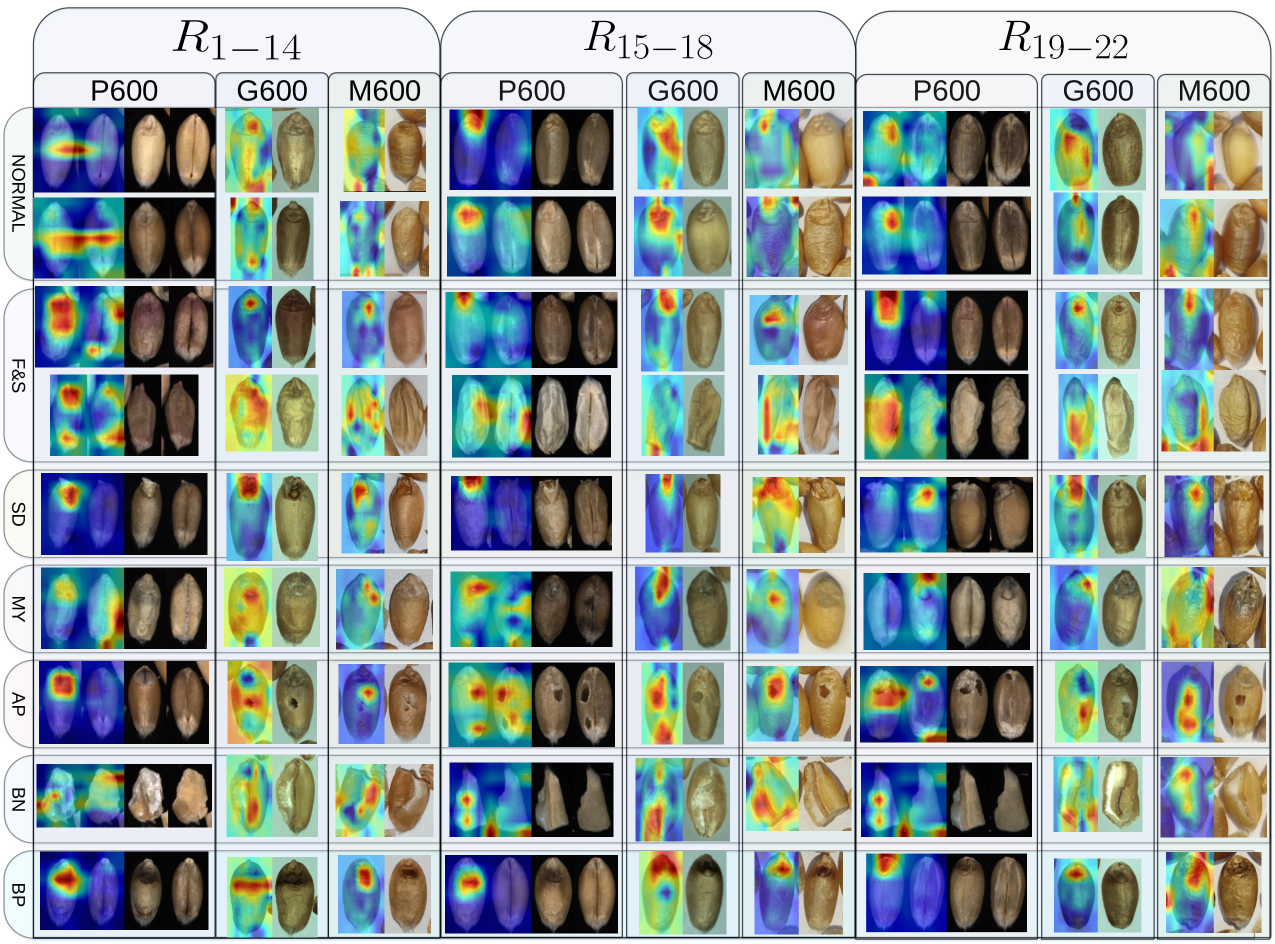

Figure XII: CAM-based visualization of single-kernel images of wheat grains.

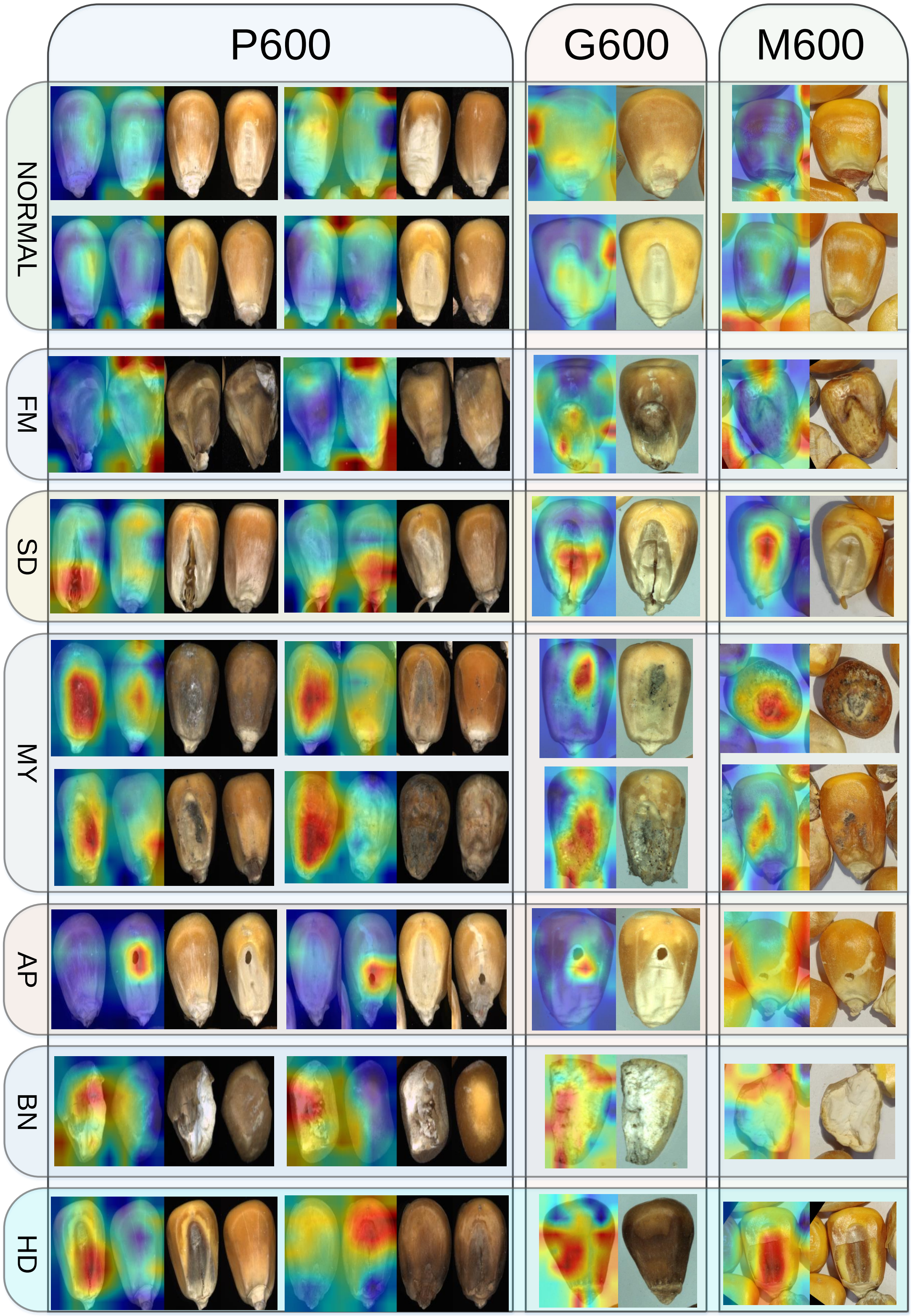

Figure XIII: CAM-based visualization of single-kernel images of maize grains.

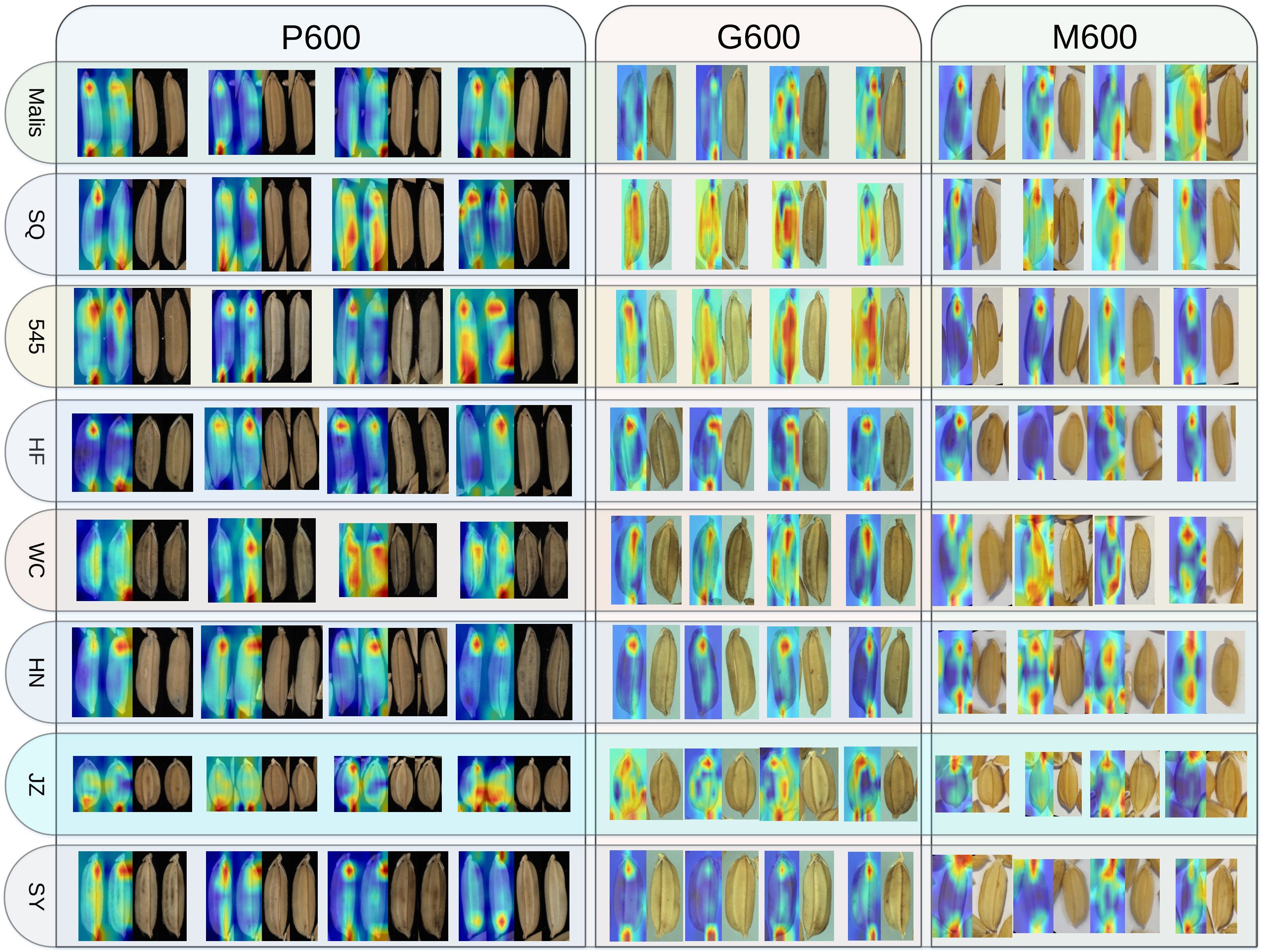

Figure XIV: CAM-based visualization of single-kernel images of rice grains.

References

[1] Anelia Angelova, Shenghuo Zhu, and Yuanqing Lin. Image segmentation for large-scale subcategory flower recognition. In *WACV*, pages 39–45. IEEE, 2013. 2

[2] Liron Bergman and Yedid Hoshen. Classification-based anomaly detection for general data. *arXiv preprint arXiv:2005.02359*, 2020. 3

[3] 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. 3

[4] David Berthelot, Nicholas Carlini, Ian Goodfellow, et al. Mixmatch: A holistic approach to semi-supervised learning. *NeurIPS*, 32, 2019. 3

[5] Konstantinos Bousmalis, George Trigeorgis, Nathan Silberman, Dilip Krishnan, and Dumitru Erhan. Domain separation networks. *NIPS*, 29, 2016. 3

[6] Rich Caruana, Yin Lou, Johannes Gehrke, Paul Koch, Marc Sturm, and Noemie Elhadad. Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. In *ACM SIGKDD*, pages 1721–1730, 2015. 3

[7] Dongliang Chang, Yifeng Ding, Jiyang Xie, et al. The devil is in the channels: Mutual-channel loss for fine-grained image classification. *TIP*, 29:4683–4695, 2020. 2

[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. 2

[9] Abhimanyu Dubey, Otkrist Gupta, Pei Guo, et al. Pairwise confusion for fine-grained visual classification. In *ECCV*, pages 70–86, 2018. 2

[10] Kevin Eykholt, Ivan Evtimov, Earlence Fernandes, et al. Robust physical-world attacks on deep learning visual classification. In *CVPR*, pages 1625–1634, 2018. 3

[11] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, et al. Domain-adversarial training of neural networks. *JMLR*, 17(1):2096–2030, 2016. 3

[12] Izhak Golan and Ran El-Yaniv. Deep anomaly detection using geometric transformations. *NIPS*, 31, 2018. 3

[13] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *CVPR*, pages 9729–9738, 2020. 3

[14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, pages 770–778, 2016. 2

[15] Dan Hendrycks, Mantas Mazeika, Saurav Kadavath, and Dawn Song. Using self-supervised learning can improve model robustness and uncertainty. *NeurIPS*, 32, 2019. 3

[16] Dan Hendrycks, Mantas Mazeika, Saurav Kadavath, and Dawn Song. Using self-supervised learning can improve model robustness and uncertainty. *NeurIPS*, 32:15663–15674, 2019. 3

[17] Ruyi Ji, Longyin Wen, Libo Zhang, et al. Attention convolutional binary neural tree for fine-grained visual categorization. In *CVPR*, pages 10468–10477, 2020. 2

[18] Ying Jin, Ximei Wang, Mingsheng Long, and Jianmin Wang. Minimum class confusion for versatile domain adaptation. In *ECCV*, pages 464–480. Springer, 2020. 3

[19] Jonathan Krause, Michael Stark, Jia Deng, and Li Fei-Fei. 3d object representations for fine-grained categorization. In *ICCV workshops*, pages 554–561, 2013. 2

[20] Chuanbin Liu, Hongtao Xie, Zheng-Jun Zha, et al. Filtration and distillation: Enhancing region attention for fine-grained visual categorization. In *AAAI*, volume 34, pages 11555–11562, 2020. 2

[21] Ze Liu, Yutong Lin, Yue Cao, et al. Swin transformer: Hierarchical vision transformer using shifted windows. In *ICCV*, pages 10012–10022, 2021. 2

[22] Mingsheng Long, Zhangjie Cao, Jianmin Wang, and Michael I Jordan. Conditional adversarial domain adaptation. In *NIPS*, pages 1647–1657, 2018. 3

[23] Xiaoxiao Ma, Jia Wu, Shan Xue, et al. A comprehensive survey on graph anomaly detection with deep learning. *TKDE*, 2021. 3

[24] Subhransu Maji, Esa Rahtu, Juho Kannala, Matthew Blaschko, and Andrea Vedaldi. Fine-grained visual classification of aircraft. *arXiv preprint arXiv:1306.5151*, 2013. 2

[25] M-E Nilsback and Andrew Zisserman. A visual vocabulary for flower classification. In *CVPR*, volume 2, pages 1447–1454. IEEE, 2006. 2

[26] Vishal M Patel, Raghuraman Gopalan, Ruonan Li, and Rama Chellappa. Visual domain adaptation: A survey of recent advances. *IEEE signal processing magazine*, 32(3):53–69, 2015. 3

[27] Clifton Phua, Vincent Lee, Kate Smith, and Ross Gayler. A comprehensive survey of data mining-based fraud detection research. *arXiv preprint arXiv:1009.6119*, 2010. 3

[28] Sanjay Purushotham, Wilka Carvalho, Tanachat Nilanon, and Yan Liu. Variational recurrent adversarial deep domain adaptation. In *ICLR*, 2016. 3

[29] Scott Reed, Zeynep Akata, Honglak Lee, and Bernt Schiele. Learning deep representations of fine-grained visual descriptions. In *CVPR*, pages 49–58, 2016. 2

[30] Lukas Ruff, Jacob R Kauffmann, Robert A Vandermeulen, et al. A unifying review of deep and shallow anomaly detection. *Proceedings of the IEEE*, 2021. 3

[31] Lukas Ruff, Robert Vandermeulen, Nico Goernitz, et al. Deep one-class classification. In *ICML*, pages 4393–4402. PMLR, 2018. 3

[32] Kuniaki Saito, Kohei Watanabe, Yoshitaka Ushiku, and Tatsuya Harada. Maximum classifier discrepancy for unsupervised domain adaptation. In *CVPR*, pages 3723–3732, 2018. 3

[33] Jihoon Tack, Sangwoo Mo, Jongheon Jeong, and Jinwoo Shin. Csi: Novelty detection via contrastive learning on distributionally shifted instances. *NeurIPS*, 33:11839–11852, 2020. 3

[34] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *JMLR*, 9(11), 2008. 3

[35] Catherine Wah, Steve Branson, Peter Welinder, Pietro Perona, and Serge Belongie. The caltech-ucsd birds-200-2011 dataset. 2011. 2

[36] Jindong Wang, Cuiling Lan, Chang Liu, Yidong Ouyang, Wenjun Zeng, and Tao Qin. Generalizing to unseen domains: A survey on domain generalization. *arXiv preprint arXiv:2103.03097*, 2021. 3

[37] Mei Wang and Weihong Deng. Deep visual domain adaptation: A survey. *Neurocomputing*, 312:135–153, 2018. 3

[38] Ze Yang, Tiange Luo, Dong Wang, Zhiqiang Hu, Jun Gao, and Liwei Wang. Learning to navigate for fine-grained classification. In *ECCV*, pages 420–435, 2018. 2

[39] Hongyi Zhang, Moustapha Cisse, Yann N Dauphin, and David Lopez-Paz. mixup: Beyond empirical risk minimization. *arXiv preprint arXiv:1710.09412*, 2017. 3

[40] Xiaopeng Zhang, Hongkai Xiong, Wengang Zhou, Weiyao Lin, and Qi Tian. Picking deep filter responses for fine-grained image recognition. In *CVPR*, pages 1134–1142, 2016. 2

[41] Yu Zhang, Xiu-Shen Wei, Jianxin Wu, et al. Weakly supervised fine-grained categorization with part-based image representation. *TIP*, 25(4):1713–1725, 2016. 2

[42] Bolei Zhou, Aditya Khosla, Agata Lapedriza, et al. Learning deep features for discriminative localization. In *CVPR*, pages 2921–2929, 2016. 4

[43] Peiqin Zhuang, Yali Wang, and Yu Qiao. Learning attentive pairwise interaction for fine-grained classification. In *AAAI*, volume 34, pages 13130–13137, 2020. 2