MANTA: A Large-Scale Multi-View and Visual-Text Anomaly Detection Dataset for Tiny Objects

Supplementary Material

Lei Fan[1] * Dongdong Fan[2] Zhiguang Hu[3] Yiwen Ding[2]
Donglin Di[4] Kai Yi[5] Maurice Pagnucco[1] Yang Song[1]
[1]UNSW Sydney [2]Gaozhe Technology [3]SCAU [4]Tsinghua University [5]University of Cambridge
https://grainnet.github.io/MANTA

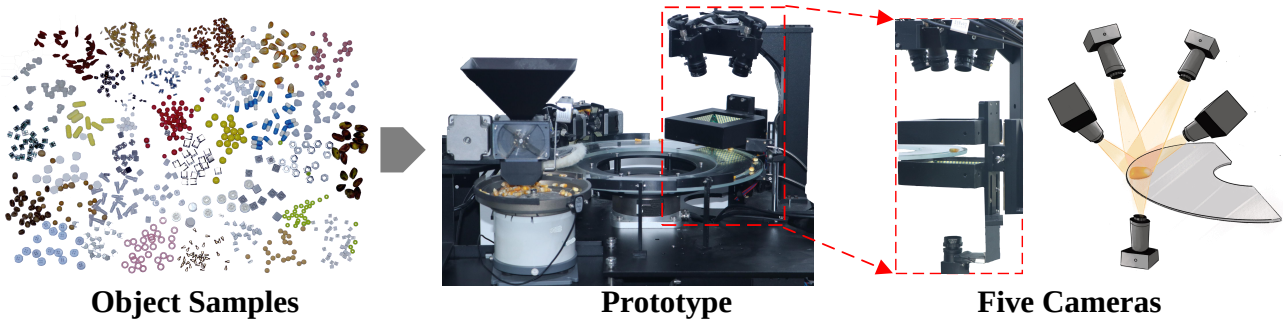


Figure I. **Multi-view images captured using the prototype.** The raw object samples were preprocessed to filter out apparent impurities before being individually fed into the device. The prototype is equipped with five high-resolution cameras. Four cameras are arranged in a quadrilateral formation, tilted downward at 30°, while one additional camera is positioned vertically beneath, pointing upward.

In this supplementary material, we provide a detailed description of MANTA and our benchmark. This document includes:

- Detailed information regarding data collection, sample visualizations, and statistical analysis of both visual and text components in **MANTA**, as provided in Section 1.
- Comprehensive benchmarking results, as provided in Section 2.
- Implementation details of advanced models, as discussed in Section 3.

1. MANTA

Data acquisition We presented detailed collected object samples and the structure of the prototype, as shown in Figure I. The raw object samples were preprocessed to remove apparent impurities before being introduced to the prototype. Each object was transferred onto a transparent plate and individually captured by the five cameras, providing comprehensive visual information.

*Corresponding author: lei.fan1@unsw.edu.au

Visual Component We showed the detailed dataset distribution, as presented in Table I, covering the normal and anomalous data across each domain and category, along with the split between training and testing sets. We also estimated the maximum bounding cuboid sizes in physical volume for each object type in the 38 categories and the storage requirements for image data in each domain. For each of the 38 categories, we randomly selected two normal and two anomalous samples, as illustrated in Figure II.

Text Component We provided detailed data distributions for the *Declarative Knowledge (DeclK)*, including both category-specific and domain-specific anomalies, as shown in Table II. We further visualized several examples of complete explicit instructions, reasoning processes, and concepts, encompassing both category-specific and domain-specific examples, as shown in Figure III.

For the *Constructivist Learning (ConsL)*, we provided detailed data distributions, including combinations of varying difficulty levels with different Normal-Normal (N-N) and Normal-Anomaly (N-A) image pairs for each category, as shown in Table III. Additionally, we showcased multiple-choice questions (MCQs) of varying difficulty levels across

different domains, along with their corresponding conclusions, as illustrated in Figure IV.

2. Detailed Results

We provided specific experimental results across multiple benchmarking experiments:

- Table IV provides detailed results corresponding to **Table 2**, using the *single-view* setting for each class. Each view of multi-view images is treated as an independent training sample. Detailed results are reported in terms of I-/P-AUROC for each category.
- Table V presents detailed results corresponding to **Table 3**, using the *multi-view* setting for each class. Multi-view images are directly used to train the models. Due to memory limitations, several advanced models were modified to adapt to the multi-view setting. Results are reported in terms of I-/P-AUROC for each category.
- Table VI shows detailed results corresponding to **Table 4**, using the *multi-class* setting. Multiple categories within the same domain are mixed, and each view is treated as a training sample. Detailed results are reported in terms of I-/P-AUROC for each category.
- Table VII provides detailed results corresponding to **Figure 9**, using *text-prompt* setting. Text information provided in the Declarative Knowledge is used to train the anomaly detection model. Results are reported in terms of I-/P-AUROC for each category.

3. Implementation Details

3.1. Single-class Models

We provided detailed experimental settings for various benchmarking models under *single-view* and *multi-view* settings. Typically, we leveraged the official code for each method to evaluate the models. The inputs were resized as 224×224 for *single-view* and 224×1120 for *multi-view*.

- RD [2]: The model was trained for 40 epochs with a batch size of 8. The Adam optimizer [7] was employed, the learning rate was set to 0.005, the Cosine Similarity loss function was utilized, and Wideresnet50 [5] was selected as the backbone.
- PatchCore [16]: The model operated with a batch size of 16. The backbone is Wideresnet50 [5], utilizing layers 2 and 3. Notably, the percentage parameter was set to 0.01 for sample selection. To prevent memory overflow, coreset operations are executed after processing every $batch\_size \times batch\_size$ samples.
- CDO [1]: The model was trained for 100 epochs with a batch size of 16. The AdamW optimizer [13] was employed with a weight decay of 0.0001. The learning rate was initialized at 4×10^{-4}, using ExponentialLR with $\gamma = 0.95$. HRNet32 [19] was utilized as the backbone.
- DMAD [11]: We used the PPDM version. The model was trained for 50 epochs with a batch size of 4. The AdamW optimizer [13] was used, and the learning rate was initialized at 0.005, using CosineAnnealingLR with $T_{\max} = 50$. The Cosine Similarity loss function was employed, and Wideresnet50 [5] was utilized as the backbone.
- SimpleNet [12]: It consists of a two-stage training process, with 20 meta-training epochs and 4 GAN-training epochs. A batch size of 16 is used. The backbone is a Wideresnet50 [5], with its feature extraction layers frozen during training. A discriminator with 2 layers and a hidden dimension of 1024 is optimized using Adam (weight decay 1×10^{-5}) [7] with a learning rate of 0.0002. The discriminator incorporates a margin threshold ($dsc\_margin = 0.15$) and Gaussian noise ($noise\_std = 0.05$) for robustness. Additionally, a pre-projector with a dimension of 1536 is optimized using Adam with a learning rate of 1×10^{-4}.

3.2. Multi-class Models

We provided detailed experimental settings for various benchmarking models under *multi-class* setting. Typically, we leveraged the official code for each method to evaluate the models. The inputs were resized as 224×224 for the *single-view* training.

- UniAD [21]: The model was trained for 50 epochs with a batch size of 32. The optimizer used was AdamW [13] with a learning rate of 4×10^{-4}, β-parameters $[0.9, 0.999]$, and a weight decay of 0.0001. The learning rate was scheduled using StepLR with a step size of 800 and a decay factor $\gamma = 0.1$. The loss function employed was FeatureMSELoss. The backbone was EfficientNet-B4 [18], utilizing layers 1, 2, 3, and 4.
- CRAD [8]: The model was trained for 20 epochs with a batch size of 8. The optimizer used was AdamW [13], with separate learning rates for different parameters: $grid\_lr = 0.1$ for trainable query parameters and $net\_lr = 0.001$ for other parameters. The learning rate scheduler was StepLR, with a step size of 40 and a decay factor $\gamma = 0.1$. The loss function employed was FeatureMSELoss. The backbone was EfficientNet-B4 [18], using layers 3 and 4.
- HGAD [20]: The backbone used is EfficientNet-b6 [18], and the flow model is a conditional-flow model [4] with 12 coupling layers and a clamping hyperparameter ($clamp\_alpha = 1.9$). Features are extracted from 3 levels. Training consists of two stages: 10 meta-epochs and 8 sub-epochs. The batch size is set to 8. The optimizer is Adam [7] with a learning rate of 2×10^{-4}. Learning rate decay is applied at epochs 50, 75, and 90 with a decay rate of 0.1. Additionally, a warming-up phase is employed with a warm-up period of 2 epochs.

3.3. Text-prompt Models

We provided detailed experimental settings for various benchmarking models under *text-prompt* setting. Typically, we leveraged the official code for PromptAD and VCP-CLIP to evaluate the models.

- WinCLIP [6]: We used an unofficial code[1]. The batch size was set to 1, and a k-shot setting with $k = 1$ was used to construct the normal reference image feature memory. The input image size was 240×240, with ViT-B-16-plus-240 [3] as the image encoder and laion400m-e31 [17] as the text encoder. For textual input, universal nouns and adjectives were combined with domain-specific nouns and adjectives to create descriptive phrases. Anomalous phrases were formatted as "category with noun" and "adjective category", while normal phrases follow the format "category without noun". All phrases are directly tokenized without any sampling or filtering.
- PromptAD [10]: The model was trained for 20 epochs with a batch size of 256 and a k-shot setting of $k = 1$. The input image size was set to 240×240, with ViT-B-16-plus-240 [3] as the image encoder and laion400m-e31 [17] as the text encoder. The optimizer used is SGD, configured with a learning rate of 0.002, momentum of 0.9, and a weight decay of 0.0005. The learning rate scheduler was CosineAnnealingLR, with $T_{\text{max}} = 20$ and $\eta_{\text{min}} = 1 \times 10^{-5}$.
- VCP-CLIP [14]: The model was trained for 2 epochs with a batch size of 32 and a k-shot setting of $k = 1$. The input image size is 518×518, with ViT-L-14-336 [3] as the image encoder and CLIP text encoder [15]. The learning rate is set to 0.00004. The text setup includes a single token and a learnable text prompt embedding with 11 layers. To prevent memory overflow, 10 normal texts and 10 anomalous texts are randomly sampled and tokenized for each run.

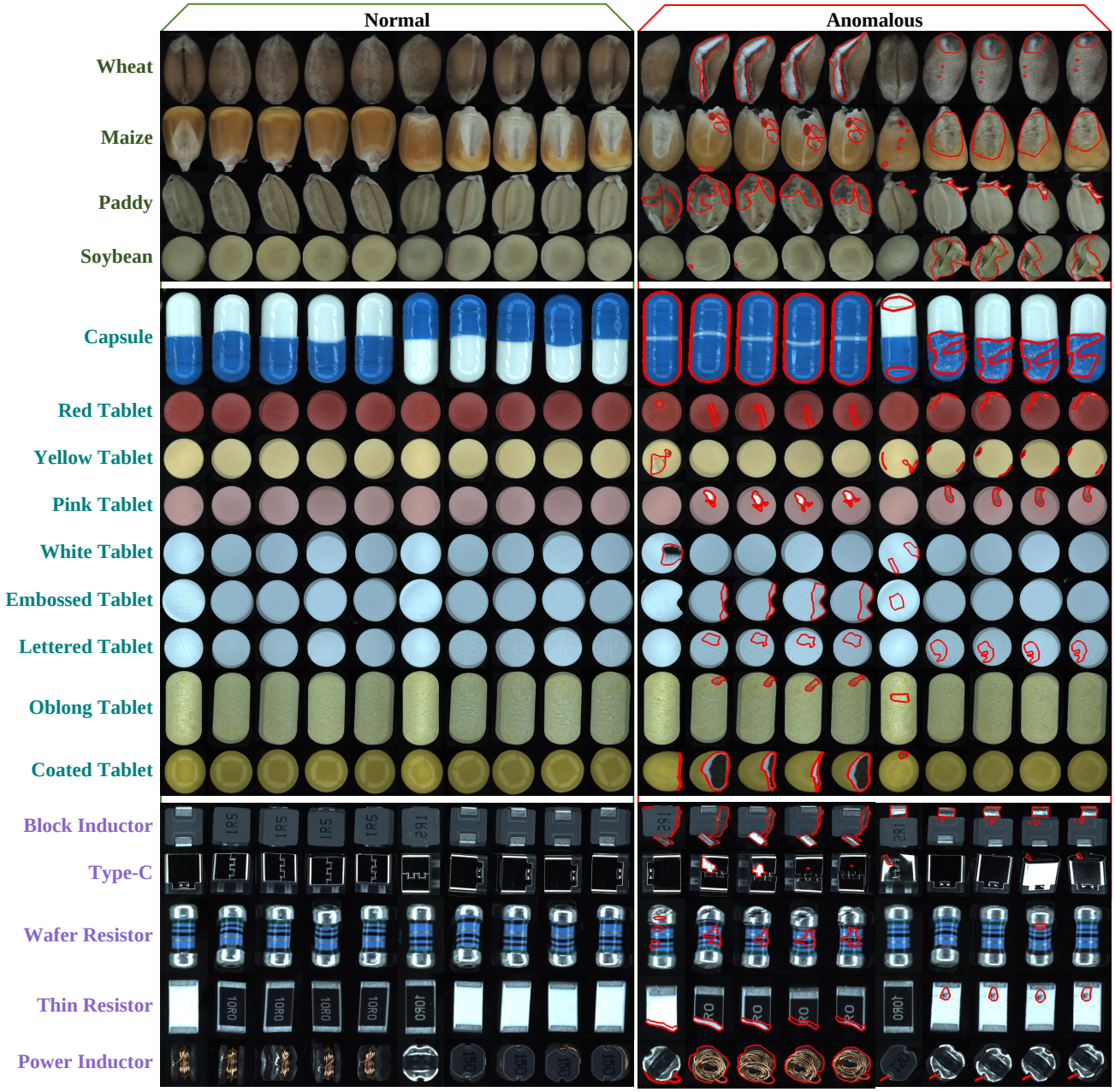
3.4. Visual-Language Model

Our baseline employed BLIP-2 [9] as the backbone. In the zero-shot setting, a specific test sample was selected, and the input data was constructed by concatenating the reference image and the test image into a single composite image. The model was then prompted to generate predictions. In the few-shot setting, one question-and-answer pair is randomly chosen and assigned to the test image. The model is trained for 11 epochs. During testing, the reference image and test image are concatenated into a composite image, and the model is prompted to generate predictions.

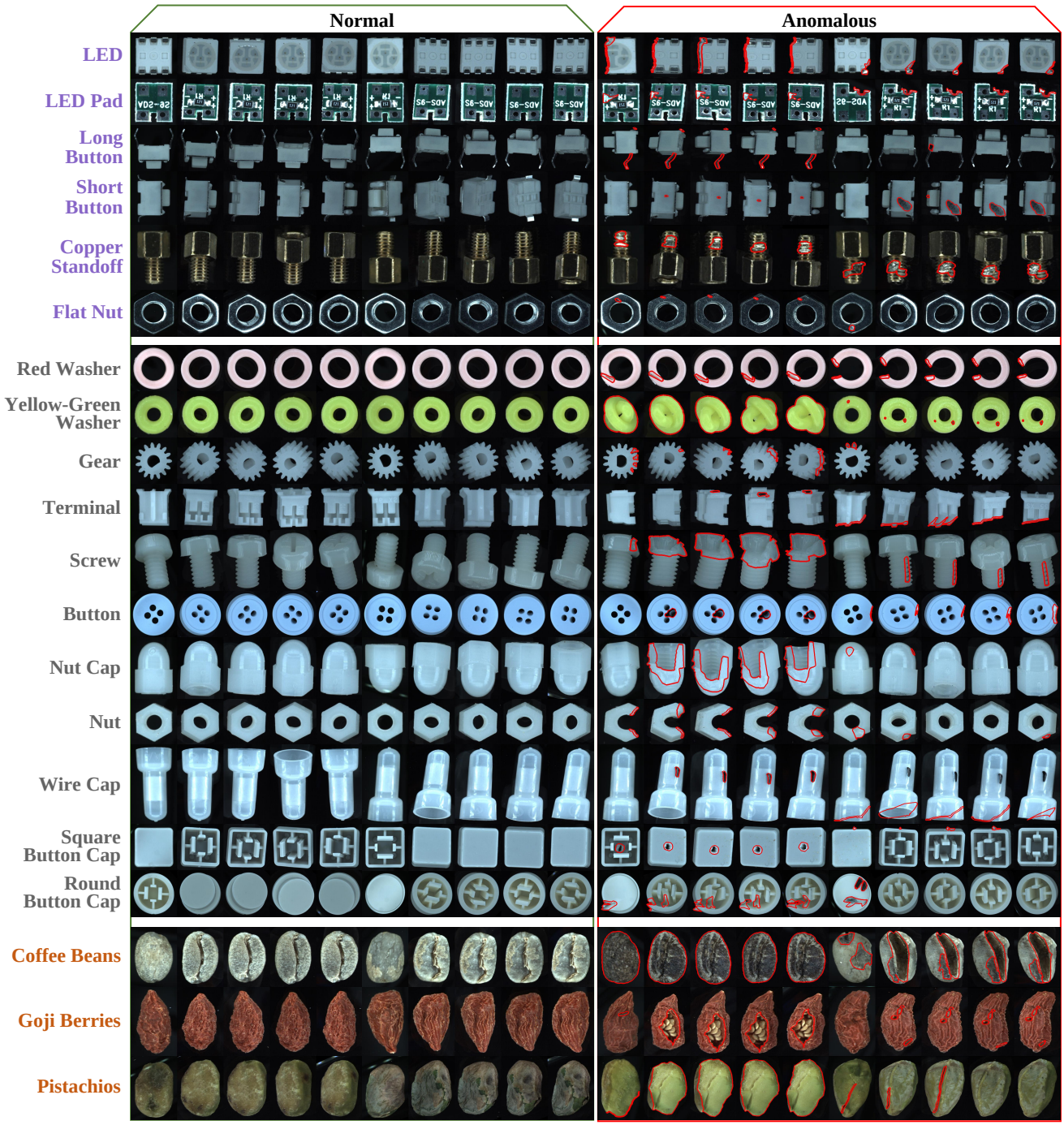
[1] https://github.com/mala-lab/WinCLIP

Table I. **Detailed data distribution of the visual component in MANTA.** The dataset comprises 137338 multi-view images spanning 38 categories from five representative domains. The required storage space for images in each domain is provided, with physical sizes estimated based on their minimum bounding cuboids. Data distribution is detailed for each category and domain.

Field	Category	Physical Size (mm^3)	Train	Test		Total	Domain Train	Domain Test		Domain Total
				Normal	Anomalous			Normal	Anomalous	
Agriculture (28.8GB)	Wheat	4×5×8	16410	3300	1650	21360	35000	10000	5000	50000
	Maize	8×10×20	9425	2800	1400	13625				
	Paddy	3×4×12	3890	1860	930	6680				
	Soybean	5×12×12	5275	2040	1020	8335				
Medicine (23.6GB)	Capsule	4×7×18	4067	786	393	5246	28579	3098	1542	33219
	Red Tablet	2×9×9	2865	1186	586	4637				
	Yellow Tablet	2×9×9	3507	102	58	3667				
	Pink Tablet	2×9×9	3624	182	91	3897				
	White Tablet	3×10×10	2382	196	98	2676				
	Embossed Tablet	3×13×13	2593	126	63	2782				
	Lettered Tablet	2×8×8	3340	138	69	3547				
	Oblong Tablet	3×9×20	2886	194	97	3177				
	Coated Tablet	3×11×11	3315	188	87	3590				
Electronics (22.4GB)	Block Inductor	2×9×10	1954	168	84	2206	22003	1370	685	24058
	Type-C	2×10×10	2178	120	60	2358				
	Wafer Resistor	2×4×8	2094	92	46	2232				
	Thin Resistor	2×4×7	2095	92	46	2233				
	Power Inductor	4×7×8	2357	140	70	2567				
	LED	2×7×7	2114	84	42	2240				
	LED Pad	1×8×9	782	166	83	1031				
	Long Button	4×9×10	1729	160	80	1969				
	Short Button	3×6×10	2057	200	100	2357				
	Copper Standoff	5×5×8	2078	70	35	2183				
	Flat Nut	5×11×12	2565	78	39	2682				
Mechanics (32.1GB)	Red Washer	3×11×11	2076	148	74	2298	22046	1726	863	24635
	Yellow-Green Washer	3×8×8	2072	144	72	2288				
	Gear	7×10×10	2068	162	81	2311				
	Terminal	7×7×9	1956	244	122	2322				
	Screw	11×11×14	2102	148	74	2324				
	Button	2×12×12	1937	178	89	2204				
	Nut Cap	10×10×12	1221	66	33	1320				
	Nut	4×8×8	2148	46	23	2217				
	Wire Cap	8×8×19	2394	220	110	2724				
	Square Button Cap	4×12×12	2041	204	102	2347				
	Round Button Cap	5×14×14	2031	166	83	2280				
Groceries (7.9GB)	Coffee Beans	7×10×14	1947	592	296	2835	3845	1054	527	5426
	Goji Berries	7×10×19	562	236	118	916				
	Pistachios	8×13×21	1336	226	113	1675				



(a) Normal and anomalous samples for 4, 9, 5 categories in the Agriculture, Medicine, and Electronics domain respectively.



(b) Normal and anomalous samples for 6, 11, 3 categories in the Electronics, Mechanics, and Groceries domain respectively.

Figure II. **Normal and anomalous samples for each of the 38 categories across five domains.** It includes two subfigures **(a)** and **(b)**. Annotated anomalous regions are highlighted with red contours, and the original images are resized to enhance visualization clarity.

Table II. **Detailed data distribution of the *Declarative Knowledge* in MANTA.** It comprises 391 category-specific and 484 domain-specific anomalies, covering 38 categories across five domains.

Domain	Category	Category-specific Anomalies	Domain-specific Anomalies	Domain Total
Agriculture	Wheat	13	130	191
	Maize	14		
	Paddy	22		
	Soybean	12		
Medicine	Capsule	28	106	170
	Red Tablet	2		
	Yellow Tablet	2		
	Pink Tablet	2		
	White Tablet	2		
	Embossed Tablet	7		
	Lettered Tablet	10		
	Oblong Tablet	7		
	Coated Tablet	4		
Electronics	Block Inductor	7	70	182
	Type-C	4		
	Wafer Resistor	10		
	Thin Resistor	10		
	Power Inductor	14		
	LED	6		
	LED Pad	25		
	Long Button	12		
	Short Button	11		
	Copper Standoff	7		
	Flat Nut	6		
Mechanics	Red Washer	15	90	216
	Yellow-Green Washer	15		
	Gear	17		
	Terminal	6		
	Screw	6		
	Button	23		
	Nut Cap	8		
	Nut	4		
	Wire Cap	10		
	Square Button Cap	11		
	Round Button Cap	11		
Groceries	Coffee Beans	12	88	116
	Goji Berries	8		
	Pistachios	8		
Total		391	484	875

"domain": *"agriculture"*,
"domain-specific": *"pest-ridden"*,

"**Reasoning**": "*caused by insect infestations that damage the kernel, leading to discoloration and decay*",

"**Concepts**":
"color": "*discolored, often yellow or brown*",
"location": *"scattered across the kernel surface*",
"area size": *"variable, from small spots to larger areas*",
"shape": "*irregular patches or lesions*",
"quantity": "*multiple, can cover significant areas*",

"domain": *"agriculture"*,
"domain-specific": *"sprouting"*,

"**Reasoning**": "*it indicates that the seed has absorbed moisture and is beginning to germinate, which is a natural developmental process under suitable conditions*",

"**Concepts**":
"color": "*green or pale green*",
"location": "*emerging from seed coat*",
"area size": "*small, linked to seed size*",
"shape": "*cylindrical or elongated*",
"quantity": "*one to several sprouts*",

"domain": *"medicine"*,
"domain-specific": *"crack"*,

"**Reasoning**": "*stress or impact causing a crack to form in the tablet*",

"**Concepts**":
"color": "*darker pink or grayish around crack*",
"location": "*surface area*",
"area size": "*varies with crack length*",
"shape": "*linear or jagged line*",
"quantity": "*single or multiple cracks*",

"domain": *"electronics"*,
"domain-specific": *"fragmented"*,

"**Reasoning**": "*mechanical stress or impact causing physical breakage*",

"**Concepts**":
"color": "*same as standard button color*",
"location": "*surface area or body*",
"area size": "*small to medium*",
"shape": "*irregular fragments*",
"quantity": "*one or more fragments*",

"domain": *"mechanics"*,
"domain-specific": *"abrasion"*,

"**Reasoning**": "*friction from repeated use or contact with rough surfaces, leading to surface wear*",

"**Concepts**":
"color": "*original color or faded areas*",
"location": "*surface of the button cap*",
"area size": "*variable, small to medium abrasion spots*",
"shape": "*flat or slightly indented areas*",
"quantity": "*single or multiple abrasions*",

"domain": *"groceries"*,
"domain-specific": *"deformation"*,

"**Reasoning**": "*genetic variation, environmental stress, or physical pressure during growth*",

"**Concepts**":
"color": "*green or pale*",
"location": "*nut inside the shell*",
"area size": "*variable, can affect part or all of the nut*",
"shape": "*misshapen or warped*",
"quantity": "*single or multiple nuts affected*",

"domain": *"agriculture"*,
"category": *"maize"*,
"category-specific": *"mildew"*,

"**Reasoning**": "*it is caused by fungal pathogens, primarily from the genus erysiphe, which thrive in humid conditions and lead to a powdery fungal growth on the seed surface*",

"**Concepts**":
"color": "*white to gray powdery coating*",
"location": "*surface of the kernel*",
"area size": "*variable, can cover small to large areas*",
"shape": "*powdery or fuzzy appearance*",
"quantity": "*variable, may cover several kernels*",

"domain": *"medicine"*,
"category": *"capsule"*,
"category-specific": *"cap-body splitting"*,

"**Reasoning**": "*improper sealing or stress during processing leading to separation of the capsule cap from the body*",

"**Concepts**":
"color": "*same as capsule body color*",
"location": "*joining area of cap and body*",
"area size": "*varies with split length*",
"shape": "*uneven or jagged split*",
"quantity": "*single or multiple splits*",

"domain": *"medicine"*,
"category": *"white tablet"*,
"category-specific": *"humidity"*,

"**Reasoning**": "*exposure to excessive moisture causing degradation or alteration in tablet surface appearance*",

"**Concepts**":
"color": "*dull or slightly yellowish*",
"location": "*surface area*",
"area size": "*small spots or patches*",
"shape": "*surface irregularities or blisters*",
"quantity": "*few to several affected areas*",

"domain": *"electronics"*,
"category": *"copper standoff"*,
"category-specific": *"damaged threads"*,

"**Reasoning**": "*impact or pressure applied during handling or assembly*",

"**Concepts**":
"color": "*same as standard color of copper standoff*",
"location": "*threaded area*",
"area size": "*small*",
"shape": "*abraded or deformed threads*",
"quantity": "*one or more threads damaged*",

"domain": *"mechanics"*,
"category": *"gear"*,
"category-specific": *"tooth loss"*,

"**Reasoning**": "*excessive wear, mechanical stress, or improper alignment leading to tooth detachment*",

"**Concepts**":
"color": "*gray or metallic*",
"location": "*edge of gear teeth*",
"area size": "*scattered or dense tooth areas*",
"shape": "*jagged or uneven edge*",
"quantity": "*single or multiple teeth*",

"domain": *"groceries"*,
"category": *"coffee beans"*,
"category-specific": *"spoilage"*,

"**Reasoning**": "*decomposition due to improper storage conditions, such as excessive moisture or heat*",

"**Concepts**":
"color": "*dark brown or black*",
"location": "*surface of the beans*",
"area size": "*variable, often patchy*",
"shape": "*soft or mushy spots*",
"quantity": "*single or multiple beans affected*",

Figure III. **Examples in Declarative Knowledge**. Both domain-specific and category-specific anomalies are shown in complete explicit instructions, reasoning, and concepts. Each concept includes five visual attributes: color, location, area size, shape, and quantity.

Table III. **Detailed data distribution of the *Constructivist Learning* in MANTA.** It includes 2000 multiple-choice questions featuring different normal and anomalous images spanning both easy and hard difficulty levels. It comprises 499 Normal-Normal (N-N, easy), 1002 Normal-Anomaly (N-A, easy), and Normal-Anomaly (N-A, hard) samples.

Domain	Category	Category				Domain			
		N-N (easy)	N-A (easy)	N-A (hard)	Total	N-N (easy)	N-A (easy)	N-A (hard)	Total
Agriculture	Wheat	33	66	33	132	100	200	100	400
	Maize	28	56	28	112				
	Paddy	19	37	19	75				
	Soybean	20	41	20	81				
Medicine	Capsule	25	52	25	102	98	202	98	398
	Red Tablet	38	76	38	152				
	Yellow Tablet	3	8	3	14				
	Pink Tablet	6	12	6	24				
	White Tablet	6	13	6	25				
	Embossed Tablet	4	8	4	16				
	Lettered Tablet	4	9	4	17				
	Oblong Tablet	6	13	6	25				
	Coated Tablet	6	11	6	23				
Electronics	Block Inductor	12	25	12	49	101	199	101	401
	Type-C	8	18	8	34				
	Wafer Resistor	7	13	7	27				
	Thin Resistor	7	13	7	27				
	Power Inductor	11	20	11	42				
	LED	6	12	6	24				
	LED Pad	12	24	12	48				
	Long Button	12	23	12	47				
	Short Button	15	30	15	60				
	Copper Standoff	5	10	5	20				
	Flat Nut	6	11	6	23				
Mechanics	Red Washer	9	17	9	35	100	201	100	401
	Yellow-Green Washer	7	17	7	31				
	Gear	9	19	9	37				
	Terminal	14	28	14	56				
	Screw	9	17	9	35				
	Button	10	21	10	41				
	Nut Cap	3	8	3	14				
	Nut	5	5	5	15				
	Wire Cap	13	26	13	52				
	Square Button Cap	11	24	11	46				
	Round Button Cap	10	19	10	39				
Groceries	Coffee Beans	57	112	57	226	100	200	100	400
	Goji Berries	22	45	22	89				
	Pistachios	21	43	21	85				
Total						499	1002	499	2000

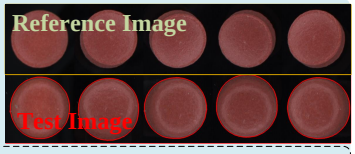
"domain": *"agriculture"*,
"category": *"wheat"*,
"difficulty": *"N-N"*,

"**Q1**": "*Is the image normal?*",
"answer": **true**,
"**Q2**": "*Is there a disease spot?*",
"answer": **false**,
"**Q3**": "*Are there any pest damages?*",
"answer": **false**,
"**Q4**": "*Is there any mold?*",
"answer": **false**,
"**Q5**": "*Is it free of shriveling?*",
"answer": **true**,
"**conclusion**": "*The wheat grains shown in the image are normal, with no anomalies such as disease spots, pest damages, mold, or shriveling.*"

"domain": *"agriculture"*,
"category": *"paddy"*,
"difficulty": *"N-N"*,

"**Q1**": "*Are there any spot anomalies in this image?*",
"answer": **false**,
"**Q2**": "*Is there any sign of mold in this image?*",
"answer": **false**,
"**Q3**": "*Is this image normal?*",
"answer": **true**,
"**Q4**": "*Is it free of pest damage in this image?*",
"answer": **true**,
"**Q5**": "*Are there any crack anomalies in this image?*",
"answer": **false**,
"**conclusion**": "*This rice image is normal, with no abnormalities observed, and no signs of spots, pests, or cracks.*"

"domain": *"medicine"*,
"category": *"red tablet"*,
"difficulty": *"N-A, easy"*,

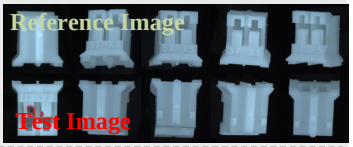


"**Q1**": "*Are there any swelling anomalies in this image?*",
"answer": **true**,
"**Q2**": "*Is there a disease spot?*",
"answer": **false**,
"**Q3**": "*Are there no signs of shriveling?*",
"answer": **true**,
"**Q4**": "*Is there no heat damage in the image?*",
"answer": **true**,
"**Q5**": "*Are there signs of mold?*",
"answer": **false**,
"**conclusion**": "*The five views show swelling abnormalities mainly concentrated on the left and right sides of views 1, 3, 4, and 5. The shape is clearly different from normal grains, and there are no signs of disease spots, shriveling, heat damage, or mold.*"

"domain": *"electronics"*,
"category": *"block inductor"*,
"difficulty": *"N-A, easy"*,

"**Q1**": "*Anything anomalous?*",
"answer": **true**,
"**Q2**": "*Is there any mutilation?*",
"answer": **true**,
"**Q3**": "*Are there any cracks?*",
"answer": **false**,
"**Q4**": "*Is it free of blemishes?*",
"answer": **false**,
"**Q5**": "*Are there any foreign objects?*",
"answer": **false**,
"**conclusion**": "*In this image, mutilations are visible in the 2nd, 3rd, 4th, and 5th views, located at the lower left and below, showing irregular gaps that indicate the integrity of this electronic component is compromised. No signs of cracks or foreign objects are present, thus these two anomalies do not exist.*"

"domain": *"mechanics"*,
"category": *"terminal"*,
"difficulty": *"N-A, hard"*,



"**Q1**": "*Does it affect the appearance quality of the terminal?*",
"answer": **true**,
"**conclusion**": "*Color spots may affect the appearance quality of the terminal, making it look unattractive.*"

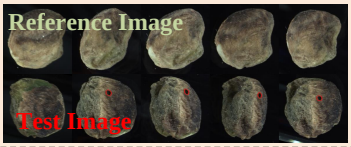
"**Q2**": "*Does it cause issues during connection of the terminal?*",
"answer": **false**,
"**conclusion**": "*Color spots do not directly affect the connection functionality of the terminal, thus not impacting its use during connection.*"

"**Q3**": "*Could it affect the durability of the terminal?*",
"answer": **false**,
"**conclusion**": "*Color spots primarily affect appearance and will not impact the durability of the terminal.*"

"**Q4**": "*Is it a defect generated during the manufacturing process?*",
"answer": **true**,
"**conclusion**": "*Color spots are typically defects that occur during the manufacturing process and may be related to materials or techniques.*"

"*Q5*": "*Is it possible that it affects the quality standards of the terminal product?*",
"answer": **true**,
"**conclusion**": "*Color spots may prevent the terminal product from meeting quality standards, thus affecting its overall compliance.*"

"domain": *"groceries"*,
"category": *"pistachios"*,
"difficulty": *"N-A, hard"*,



"**Q1**": "*Is the eating quality affected?*",
"answer": **true**,
"**conclusion**": "*The presence of worm infestation could lead to spoilage of the pistachios, directly impacting their eating quality.*"

"**Q2**": "*Does it reduce the nutrient content of pistachios?*",
"answer": **true**,
"**conclusion**": "*Worm infestation may lead to the loss of the nutritional content of the pistachios, thus affecting their overall value.*"

"**Q3**": "*Were effective pest control measures taken during storage?*",
"answer": **false**,
"**conclusion**": "*The presence of worm infestation indicates that effective pest control measures were not taken during storage, leading to product damage.*"

"**Q4**": "*Does it affect the market value of pistachios?*",
"answer": **true**,
"**conclusion**": "*Worm infestation typically significantly decreases the market value of the product as it compromises its integrity.*"

"**Q5**": "*Are storage conditions improper?*",
"answer": **true**,
"**conclusion**": "*Worm infestation is usually associated with humidity or improper storage conditions, which may lead to pest occurrence.*"

Figure IV. **Examples in Constructivist Learning**. Each multiple-choice question comprises a pair of images, five questions, and a conclusion. The reference image, which is normal, serves as the image prompt, while the questions are designed to be answered based on the test image. For easy-level questions, a total conclusion is provided. For hard-level questions, a detailed conclusion is provided for each question. N-A denotes Normal-Anomaly image pairs.

Table IV. **Detailed results for Table 2, *single-view* setting for each class.** Models are trained using single-view images and reported results in both *view*-eval and *object*-eval (predictions from five views of an object). All results are presented as I-/P-AUROC (%).

Domain	Category	RD [2]		PatchCore [16]		CDO [1]		DMAD [11]		SimpleNet [12]	
		view-eval	*object*-eval	*view*-eval	*object*-eval	*view*-eval	*object*-eval	*view*-eval	*object*-eval	*view*-eval	*object*-eval
Agriculture	Wheat	84.9/84.8	89.2/80.8	96.6/96.9	98.3/96.9	93.1/96.5	93.6/96.2	78.4/85.8	85.5/85.0	85.0/87.1	90.5/88.6
	Maize	82.7/85.8	84.7/88.4	86.9/92.8	88.3/92.8	85.9/93.1	86.8/93.2	78.9/83.8	83.4/85.9	81.6/82.4	88.3/83.6
	Paddy	85.3/86.2	87.6/86.1	93.3/88.8	95.1/88.7	87.1/80.1	90.2/79.8	90.2/81.8	91.3/86.5	86.6/79.3	92.8/82.5
	Soybean	85.8/83.9	87.5/85.9	95.3/93.6	95.2/93.5	94.0/92.7	94.6/92.4	90.4/88.3	87.4/83.3	92.0/84.6	93.1/82.4
	Average	84.7/85.2	87.2/85.3	93.0/93.0	94.2/93.0	90.0/90.6	91.3/90.4	84.5/84.9	86.9/85.2	86.3/83.3	91.2/84.3
Medicine	Capsule	96.8/95.7	97.9/94.6	99.0/89.2	98.9/89.2	98.9/90.6	98.0/90.5	96.9/96.2	97.9/94.2	98.7/80.7	98.9/88.1
	Red Tablet	86.5/92.4	92.3/94.4	90.4/90.0	90.1/91.9	88.3/81.0	87.7/81.7	78.5/89.9	88.9/89.5	93.0/77.7	98.1/79.0
	Yellow Tablet	85.9/89.8	89.9/92.8	98.2/99.1	99.4/99.1	98.0/98.9	98.5/98.9	85.6/93.1	89.5/89.3	96.0/96.3	96.8/95.0
	Pink Tablet	85.0/92.9	89.8/93.6	97.3/99.3	98.9/99.3	97.1/98.7	99.2/98.7	84.7/95.0	88.5/93.0	95.6/96.0	98.5/96.6
	White Tablet	85.0/94.3	93.9/94.5	97.3/98.8	98.6/98.9	97.2/98.7	98.9/98.7	86.4/95.3	88.3/90.5	93.0/91.9	97.9/92.4
	Embossed Tablet	83.0/95.2	90.2/93.9	96.9/98.0	97.7/98.2	94.6/96.6	93.8/96.7	81.3/93.7	89.1/89.5	85.7/85.1	90.6/84.3
	Lettered Tablet	79.7/96.3	89.5/94.3	95.3/98.7	97.6/98.7	93.9/98.0	94.3/98.0	77.6/90.5	91.6/90.6	80.2/77.7	91.9/81.0
	Oblong Tablet	85.2/94.6	90.3/92.9	94.7/97.8	95.8/97.8	90.6/96.8	86.7/96.9	78.9/89.5	90.3/89.3	80.4/80.5	80.6/79.1
	Coated Tablet	90.9/94.7	96.8/95.7	99.2/99.8	99.8/99.8	98.4/99.5	99.4/99.6	91.3/96.7	97.0/94.6	97.1/99.3	99.0/99.2
	Average	86.4/94.0	92.3/94.1	96.5/96.7	97.4/97.0	95.2/95.4	95.2/95.5	84.6/93.3	91.2/91.2	91.1/87.2	94.7/88.3
Electronics	Block Inductor	83.1/92.4	84.7/93.4	94.1/99.0	93.3/98.9	87.9/98.3	83.5/98.2	88.5/93.2	88.9/93.7	89.2/92.8	93.8/93.2
	Type-C	88.4/94.8	90.5/94.8	98.4/98.8	99.4/98.8	96.6/98.9	98.6/98.9	90.3/93.0	90.8/92.1	92.5/91.4	96.6/93.0
	Wafer Resistor	85.4/93.9	87.6/92.0	96.4/99.4	96.2/99.4	95.7/99.0	94.2/99.0	87.7/91.9	88.5/91.6	90.6/96.1	93.2/96.4
	Thin Resistor	84.9/94.0	91.9/94.0	98.8/97.9	99.9/98.0	96.7/97.9	95.8/98.0	85.5/93.9	92.7/91.7	91.0/90.8	93.7/91.0
	Power Inductor	84.8/89.8	89.7/93.2	91.2/97.3	89.4/97.2	86.9/96.9	85.2/97.0	83.8/87.0	89.6/86.3	84.7/87.6	85.9/91.7
	LED	87.1/94.2	91.0/93.2	99.2/99.5	99.0/99.5	97.7/99.4	96.5/99.4	88.0/90.4	92.8/92.3	94.0/94.3	98.6/95.5
	LED Pad	87.9/95.2	89.7/92.5	99.3/98.3	99.1/98.3	96.6/98.6	90.2/98.6	79.2/89.4	90.6/92.7	94.2/90.4	96.0/92.3
	Long Button	94.5/95.5	92.0/95.5	98.6/98.6	97.4/98.7	97.2/98.8	94.6/98.6	94.5/95.7	92.7/96.1	92.7/92.2	92.3/90.4
	Short Button	81.0/94.8	88.4/93.3	97.4/99.4	98.1/99.4	95.3/99.6	93.9/99.4	88.2/93.1	89.0/91.9	86.3/88.2	90.9/86.7
	Copper Standoff	89.3/92.6	88.4/92.6	99.0/99.0	99.4/99.1	96.8/98.6	99.5/98.7	90.8/93.2	87.8/93.4	87.8/84.1	89.6/85.5
	Flat Nut	86.3/91.6	87.1/92.6	95.7/99.3	95.7/99.3	87.5/99.0	79.3/98.8	84.9/91.8	87.9/91.8	83.0/84.9	87.1/87.7
	Average	86.6/93.5	89.2/93.4	97.1/98.8	97.0/98.8	94.1/98.6	91.9/98.6	87.4/92.1	90.1/92.2	89.6/90.3	92.5/91.2
Mechanics	Red Washer	79.1/93.0	89.3/92.1	98.7/99.4	98.4/99.4	96.5/99.2	94.2/99.3	83.8/91.3	85.2/91.3	95.3/96.6	94.9/95.8
	Yellow-Green Washer	85.8/94.7	89.1/93.7	94.0/95.2	95.1/95.3	90.1/94.7	89.9/94.3	83.6/92.5	89.3/91.6	88.7/88.8	91.3/87.5
	Gear	88.9/94.2	86.1/90.3	96.8/99.3	97.8/99.3	91.3/99.3	88.0/99.3	77.2/89.3	86.3/89.1	88.8/90.4	92.1/89.9
	Terminal	84.7/88.2	87.8/89.8	96.8/99.0	97.8/99.0	89.6/98.6	85.5/98.7	82.2/89.4	85.2/89.7	80.1/82.0	79.3/79.8
	Screw	89.9/92.8	88.9/89.9	92.1/98.1	96.3/98.1	81.4/96.9	82.0/96.9	83.7/89.1	78.8/87.4	77.8/78.6	87.0/77.0
	Button	77.0/89.9	82.7/91.8	94.0/99.6	91.5/99.6	92.4/99.6	89.7/99.6	72.7/88.2	81.1/89.7	86.5/94.0	86.2/93.2
	Nut Cap	68.4/89.2	89.5/91.2	91.6/98.1	93.2/98.1	84.9/97.8	91.2/97.9	79.6/88.7	83.4/89.4	75.4/83.2	88.2/82.6
	Nut	58.6/88.3	62.4/91.5	96.8/99.3	96.6/99.3	93.5/99.1	91.9/99.0	70.4/87.6	82.5/87.8	84.7/91.1	85.2/90.9
	Wire Cap	78.8/92.8	85.0/92.8	95.7/98.7	96.2/98.7	88.8/98.8	88.1/98.8	78.9/88.2	84.4/90.0	86.6/89.9	87.1/89.5
	Square Button Cap	91.1/94.1	96.9/94.3	98.2/99.4	98.7/99.4	96.3/99.4	96.9/99.4	90.0/95.7	94.6/95.7	92.0/90.5	98.0/94.6
	Round Button Cap	88.6/94.0	84.9/92.0	99.2/99.5	99.3/99.5	96.7/99.5	95.6/99.5	80.3/93.6	87.6/92.4	94.5/90.4	96.7/93.0
	Average	81.0/91.9	85.7/91.8	95.8/98.7	96.4/98.7	91.1/98.4	90.3/98.4	80.2/90.3	85.3/90.4	86.4/88.7	89.6/88.5
Groceries	Coffee Beans	70.3/82.2	74.7/83.2	85.8/90.9	90.1/90.9	89.7/91.0	90.8/90.8	78.1/89.7	77.2/88.2	94.4/91.3	97.4/94.1
	Goji Berries	72.0/86.4	73.5/86.4	87.8/95.7	92.4/95.7	85.6/96.5	88.9/96.3	76.1/86.6	78.7/87.9	78.9/87.0	83.2/87.0
	Pistachios	71.4/87.3	72.4/86.4	85.0/87.8	89.9/87.9	80.9/83.6	83.4/80.8	71.2/86.3	75.0/86.2	73.1/69.3	79.0/73.8
	Average	71.2/85.3	73.5/85.3	86.2/91.5	90.8/91.5	85.4/90.4	87.7/89.3	75.1/87.5	77.0/87.4	82.2/82.5	86.5/84.9
Total Average		82.0/90.0	85.6/90.0	93.7/95.7	95.2/95.8	91.2/94.7	91.3/94.4	82.6/89.6	86.1/89.7	87.1/86.4	90.9/87.5

Table V. **Detailed results for Table 3, *multi-view* setting for each class.** Models are trained using multi-view images, and all results are presented as I-/P-AUROC (%).

Domain	Category	RD [2]		PatchCore [16]		CDO [1]		DMAD [11]	
		I-AUROC	P-AUROC	I-AUROC	P-AUROC	I-AUROC	P-AUROC	I-AUROC	P-AUROC
Agriculture	Wheat	91.4	95.2	98.3	97.2	93.1	95.8	92.7	95.7
	Maize	79.3	91.4	88.0	92.9	86.9	92.3	80.7	90.6
	Paddy	88.2	79.3	94.9	88.3	88.4	82.1	90.6	80.3
	Soybean	88.8	92.0	95.1	93.6	94.5	91.4	89.9	92.2
	Average	86.9	89.5	94.1	93.0	90.7	90.4	88.5	89.7
Medicine	Capsule	98.7	90.3	98.6	88.5	98.1	90.7	98.5	90.1
	Red Tablet	80.8	84.6	86.1	86.8	89.6	82.0	84.3	83.9
	Yellow Tablet	99.0	98.9	99.1	99.0	98.4	99.0	99.4	98.6
	Pink Tablet	98.6	99.1	98.9	99.2	98.9	98.7	98.5	98.8
	White Tablet	97.6	98.7	98.7	98.7	98.9	98.7	97.2	98.6
	Embossed Tablet	97.8	98.6	97.6	98.0	94.6	96.5	93.7	98.1
	Lettered Tablet	85.5	97.3	97.4	98.6	95.7	97.9	95.2	97.7
	Oblong Tablet	91.0	97.2	94.9	97.5	84.8	96.2	89.1	96.6
	Coated Tablet	99.7	99.8	99.7	99.8	99.2	99.5	99.9	99.7
	Average	94.3	96.1	96.8	96.2	95.4	95.5	95.1	95.8
Electronics	Block Inductor	90.7	98.0	93.1	98.8	84.6	98.2	88.6	97.8
	Type-C	98.0	99.0	99.3	98.8	98.5	98.9	97.6	98.6
	Wafer Resistor	94.3	99.2	95.8	99.4	93.4	99.1	95.7	99.2
	Thin Resistor	98.3	97.6	99.7	97.8	96.6	98.0	99.5	98.1
	Power Inductor	83.2	96.3	88.8	97.2	85.6	97.1	84.8	96.8
	LED	99.0	99.5	98.2	99.4	97.1	99.5	98.2	99.3
	LED Pad	97.4	98.7	99.2	98.2	93.6	98.4	98.4	98.5
	Long Button	96.7	98.4	97.2	98.3	95.8	98.5	96.4	98.0
	Short Button	94.7	98.8	97.9	99.4	93.8	99.5	94.1	99.1
	Copper Standoff	98.4	98.7	99.7	99.0	99.6	98.6	96.8	97.2
	Flat Nut	92.8	99.3	95.6	99.3	76.9	98.6	86.4	98.5
	Average	94.9	98.5	96.8	98.7	92.3	98.6	94.2	98.3
Mechanics	Red Washer	99.0	99.4	98.7	99.3	96.2	99.2	95.2	99.1
	Yellow-Green Washer	94.5	95.2	94.5	94.8	91.5	94.8	95.0	95.7
	Gear	97.9	99.4	97.8	99.2	89.7	99.3	94.4	99.2
	Terminal	95.9	99.1	97.6	99.0	87.8	98.5	92.4	98.9
	Screw	92.3	97.3	96.0	97.8	84.8	96.9	88.1	97.2
	Button	91.3	99.5	91.8	99.5	89.3	99.6	87.6	98.7
	Nut Cap	96.5	98.3	94.3	98.1	92.4	97.8	94.4	96.6
	Nut	99.4	99.2	97.6	99.2	89.7	99.0	94.9	98.5
	Wire Cap	93.4	98.9	96.4	98.7	90.3	98.9	91.3	98.9
	Square Button Cap	98.5	99.3	98.4	99.4	96.9	99.4	96.3	99.0
	Round Button Cap	97.9	99.5	99.2	99.5	95.5	99.5	96.8	99.3
	Average	96.1	98.7	96.6	98.6	91.3	98.4	93.3	98.3
Groceries	Coffee Beans	72.6	85.9	90.0	91.3	91.4	91.5	81.2	87.3
	Goji Berries	91.9	93.8	92.7	95.9	89.2	96.3	91.2	93.9
	Pistachios	78.9	82.1	90.3	88.2	82.7	80.8	67.7	77.7
	Average	81.1	87.3	91.0	91.8	87.8	89.5	80.0	86.3
Total Average		90.6	94.0	95.0	95.7	91.5	94.5	90.2	93.7

Table VI. **Detailed results for Table 4, *multi-class* setting.** Models are trained using single-view mixed data across all categories within each domain, and all results are presented as I-/P-AUROC (%).

Domain	Category	UniAD [21]		CRAD [8]		HGAD [20]	
		I-AUROC	P-AUROC	I-AUROC	P-AUROC	I-AUROC	P-AUROC
Agriculture	Wheat	76.2	91.6	82.8	89.3	86.3	93.3
	Maize	63.7	81.0	83.5	88.1	81.5	90.1
	Paddy	67.2	72.4	85.3	79.8	86.5	80.2
	Soybean	74.8	82.7	89.3	88.4	87.7	90.9
	Average	70.5	81.9	85.2	86.4	85.5	88.6
Medicine	Capsule	98.7	88.3	98.4	90.7	96.0	87.5
	Red Tablet	71.2	71.1	84.7	84.6	89.6	84.2
	Yellow Tablet	95.4	98.0	96.3	96.1	94.1	97.7
	Pink Tablet	95.3	98.6	97.8	96.9	93.2	97.5
	White Tablet	94.4	96.2	95.0	96.5	93.3	97.3
	Embossed Tablet	89.3	97.0	94.1	97.8	89.4	95.6
	Lettered Tablet	91.1	96.7	92.4	97.3	97.7	94.4
	Oblong Tablet	87.7	94.8	90.8	94.3	82.5	89.9
	Coated Tablet	98.6	99.0	98.4	97.3	97.5	99.1
	Average	91.3	93.3	94.2	94.6	92.6	93.7
Electronics	Block Inductor	90.0	98.0	94.6	97.1	90.0	94.9
	Type-C	96.2	98.0	94.1	96.2	94.0	96.3
	Wafer Resistor	94.8	98.9	94.3	96.4	88.8	94.1
	Thin Resistor	95.3	98.4	95.5	96.3	96.2	96.6
	Power Inductor	83.3	95.2	96.0	97.6	86.0	89.0
	LED	97.9	99.3	95.7	95.2	97.4	98.7
	LED Pad	97.2	96.7	94.7	95.8	87.5	87.6
	Long Button	95.5	95.9	95.8	97.6	93.9	95.9
	Short Button	94.9	97.5	93.4	95.4	91.2	97.4
	Copper Standoff	98.4	98.7	94.4	97.0	95.2	95.9
	Flat Nut	86.3	97.8	91.3	97.4	80.7	90.4
	Average	93.6	97.7	94.5	96.5	91.0	94.3
Mechanics	Red Washer	96.8	99.2	95.7	98.4	93.3	98.1
	Yellow-Green Washer	91.4	95.0	94.2	95.6	86.8	90.2
	Gear	92.2	95.5	96.4	92.3	84.2	93.0
	Terminal	89.5	97.3	94.6	94.3	88.2	97.1
	Screw	80.0	91.8	92.5	94.1	68.7	85.8
	Button	86.3	99.3	87.7	95.1	84.9	98.8
	Nut Cap	87.0	96.0	84.0	98.0	70.6	92.2
	Nut	92.4	98.8	92.9	99.6	75.4	93.7
	Wire Cap	90.3	98.5	93.5	99.4	84.1	96.9
	Square Button Cap	95.1	98.9	95.4	99.4	92.2	98.3
	Round Button Cap	91.2	98.9	94.9	99.5	89.9	98.4
	Average	90.2	97.2	92.9	96.9	83.5	94.8
Groceries	Coffee Beans	63.0	67.7	78.8	85.3	88.8	68.8
	Goji Berries	74.1	92.0	78.4	87.5	73.8	87.4
	Pistachios	63.0	70.8	71.0	77.8	61.8	58.9
	Average	66.7	76.8	76.1	83.5	74.8	71.7
Total Average		82.4	89.4	88.6	91.6	85.5	88.6

Table VII. **Detailed results for Figure 9, *text-prompt* setting in one-shot learning.** Models are trained using text data from *DeclK*, and all results are presented as I-/P-AUROC (%).

Domain	Category	WinCLIP [6]		PromptAD [10]		VCP-CLIP [14]	
		I-AUROC	P-AUROC	I-AUROC	P-AUROC	I-AUROC	P-AUROC
Agriculture	Wheat	75.0	89.5	76.8	90.1	62.9	74.4
	Maize	66.2	88.1	75.7	87.9	78.0	68.0
	Paddy	73.6	73.4	74.4	81.6	69.5	58.2
	Soybean	92.0	91.8	90.2	90.6	74.7	77.1
	Average	76.7	85.7	79.3	87.6	71.3	69.4
Medicine	Capsule	83.9	83.7	86.3	83.0	65.5	56.8
	Red Tablet	72.1	72.1	53.4	69.3	69.9	64.0
	Yellow Tablet	99.8	98.7	96.5	98.9	66.2	92.9
	Pink Tablet	96.7	93.9	86.7	95.9	78.5	85.0
	White Tablet	93.9	98.8	85.3	95.9	74.4	78.6
	Embossed Tablet	62.5	95.0	43.1	87.1	58.0	83.8
	Lettered Tablet	91.8	97.4	73.7	96.3	72.7	71.6
	Oblong Tablet	84.9	96.0	55.3	95.5	75.2	73.0
	Coated Tablet	95.7	99.7	93.4	98.9	39.7	70.7
	Average	86.8	92.8	74.8	91.2	66.7	75.2
Electronics	Block Inductor	67.4	90.4	58.4	87.9	74.6	68.6
	Type-C	67.9	92.7	77.5	92.3	57.0	86.8
	Wafer Resistor	79.9	97.5	80.9	96.6	53.0	81.8
	Thin Resistor	56.5	93.0	79.8	89.3	72.5	82.3
	Power Inductor	56.9	85.9	62.0	76.4	57.6	77.4
	LED	82.9	97.1	82.5	85.4	73.4	73.7
	LED Pad	37.6	85.2	71.2	88.8	67.4	82.9
	Long Button	76.3	91.7	84.6	94.7	82.0	62.9
	Short Button	90.1	98.2	76.5	94.2	66.4	69.9
	Copper Standoff	82.9	96.6	88.9	96.1	72.8	87.0
	Flat Nut	85.8	98.3	95.0	95.8	80.2	77.4
	Average	71.3	93.3	77.9	90.7	68.8	77.3
Mechanics	Red Washer	96.7	99.1	90.6	96.0	88.9	82.7
	Yellow-Green Washer	82.6	89.1	76.7	83.8	75.6	82.8
	Gear	84.3	95.4	62.7	88.2	45.2	32.7
	Terminal	62.0	88.7	80.0	89.9	72.9	78.5
	Screw	75.7	83.8	51.5	77.0	76.3	73.9
	Button	71.9	96.4	72.3	84.8	54.6	70.8
	Nut Cap	64.3	92.4	72.0	97.2	70.6	87.7
	Nut	70.8	98.1	98.6	99.4	79.9	91.3
	Wire Cap	66.6	86.2	57.6	85.1	65.5	83.8
	Square Button Cap	73.3	96.2	67.3	83.3	58.3	83.0
	Round Button Cap	83.9	96.3	72.8	78.3	50.4	70.3
	Average	75.6	92.9	72.9	87.5	67.1	76.1
Groceries	Coffee Beans	65.3	81.0	49.4	80.0	64.9	61.3
	Goji Berries	80.0	94.5	88.8	92.6	61.3	82.1
	Pistachios	64.3	80.2	72.3	83.7	70.2	74.5
	Average	69.9	85.2	70.2	85.4	65.5	72.6
Total Average		76.1	90.0	75.0	88.5	67.9	74.1

References

[1] Yunkang Cao, Xiaohao Xu, Zhaoge Liu, and Weiming Shen. Collaborative discrepancy optimization for reliable image anomaly localization. *IEEE Transactions on Industrial Informatics*, 19(11):10674–10683, 2023. 2, 11, 12

[2] Hanqiu Deng and Xingyu Li. Anomaly detection via reverse distillation from one-class embedding. In *CVPR*, pages 9737–9746, 2022. 2, 11, 12

[3] Alexey Dosovitskiy. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020. 3

[4] Denis Gudovskiy, Shun Ishizaka, and Kazuki Kozuka. Cflow-ad: Real-time unsupervised anomaly detection with localization via conditional normalizing flows. In *WACV*, pages 98–107, 2022. 2

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

[6] Jongheon Jeong, Yang Zou, Taewan Kim, et al. Winclip: Zero-/few-shot anomaly classification and segmentation. In *CVPR*, pages 19606–19616, 2023. 3, 14

[7] Diederik P Kingma. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 2

[8] Joo Chan Lee, Taejune Kim, Eunbyung Park, Simon S. Woo, and Jong Hwan Ko. Continuous memory representation for anomaly detection. *ECCV*, 2024. 2, 13

[9] Junnan Li, Dongxu Li, Silvio Savarese, and Steven Hoi. Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models. In *ICML*, pages 19730–19742. PMLR, 2023. 3

[10] Xiaofan Li, Zhizhong Zhang, Xin Tan, Chengwei Chen, Yanyun Qu, Yuan Xie, and Lizhuang Ma. Promptad: Learning prompts with only normal samples for few-shot anomaly detection. In *CVPR*, pages 16838–16848, 2024. 3, 14

[11] Wenrui Liu, Hong Chang, Bingpeng Ma, Shiguang Shan, and Xilin Chen. Diversity-measurable anomaly detection. In *CVPR*, pages 12147–12156, 2023. 2, 11, 12

[12] Zhikang Liu, Yiming Zhou, Yuansheng Xu, and Zilei Wang. Simplenet: A simple network for image anomaly detection and localization. In *CVPR*, pages 20402–20411, 2023. 2, 11

[13] I Loshchilov. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017. 2

[14] Zhen Qu, Xian Tao, Mukesh Prasad, Fei Shen, Zhengtao Zhang, Xinyi Gong, and Guiguang Ding. Vcp-clip: A visual context prompting model for zero-shot anomaly segmentation. *ECCV*, 2024. 3, 14

[15] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *ICML*, pages 8748–8763. PMLR, 2021. 3

[16] Karsten Roth, Latha Pemula, Joaquin Zepeda, Bernhard Schölkopf, Thomas Brox, and Peter Gehler. Towards total recall in industrial anomaly detection. In *CVPR*, pages 14318–14328, 2022. 2, 11, 12

[17] Christoph Schuhmann, Richard Vencu, Romain Beaumont, Robert Kaczmarczyk, Clayton Mullis, Aarush Katta, Theo Coombes, Jenia Jitsev, and Aran Komatsuzaki. Laion-400m: Open dataset of clip-filtered 400 million image-text pairs. *arXiv preprint arXiv:2111.02114*, 2021. 3

[18] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *ICML*, pages 6105–6114. PMLR, 2019. 2

[19] Jingdong Wang, Ke Sun, Tianheng Cheng, Borui Jiang, Chaorui Deng, Yang Zhao, Dong Liu, Yadong Mu, Mingkui Tan, Xinggang Wang, et al. Deep high-resolution representation learning for visual recognition. *PAMI*, 43(10):3349–3364, 2020. 2

[20] Xincheng Yao, Ruoqi Li, Zefeng Qian, Lu Wang, and Chongyang Zhang. Hierarchical gaussian mixture normalizing flows modeling for unified anomaly detection. In *ECCV*, 2024. 2, 13

[21] Zhiyuan You, Lei Cui, Yujun Shen, Kai Yang, Xin Lu, Yu Zheng, and Xinyi Le. A unified model for multi-class anomaly detection. *NeurIPS*, 35:4571–4584, 2022. 2, 13