Unexplored Faces of Robustness and Out-of-Distribution: Covariate Shifts in Environment and Sensor Domains

Supplementary Material

Eunsu Baek Keondo Park Jiyoon Kim Hyung-sin Kim
Seoul National University
{beshu9407, gundo0102, iamkjy, hyungkim}@snu.ac.kr

1. Details on *ES-Studio* and *ImageNet-ES* Implementations (Sec. 4)

This section provides details on how *ES-Studio* is built and *ImageNet-ES* is collected in *ES-Studio*.

1.1. *ES-Studio* Setup

ES-Studio is established with the primary objective of ensuring the reproducibility of our proposed dataset while minimizing external factors, focusing specifically on light conditions and camera sensors. As illustrated in Figure 1, *ES-Studio* is designed as a completely dark room with dimensions of (1.5 m $\times$ 1.5 m $\times$ 2 m), equipped with four key components (screen, camera, ceiling lamps and desktop).

In terms of the dark room setup, all sides are covered with blackout fabric to effectively block out any external light. Within the dark room, a desk with a height of 267 mm is positioned at the front, featuring the placement of a large screen (Component 1) on top and a desktop computer (Component 4) below. To prevent light reflection from the desk, it is covered with blackout fabric, extending to the floor. To prevent any image distortion, careful attention is given to the height of the camera (Component 3), ensuring it is located at a distance of 1 m from the large screen (Component 1) in a straight line. Light is controlled by two ceiling lamps (Component 2), strategically positioned at the midpoint between the large screen and the camera lens. The entire setup aims to maintain consistency and accuracy in the captured images. Additionally, to address thermal issues and minimize errors and delays during data collection, ventilation outlets are installed.

Finally, the detailed specification of each component is as follows:

- **Screen (Component 1):** The screen utilized in *ES-Studio* is an OLED TV with the model named 'LG OLED55B3FNA,' featuring a 55-inch display and a high resolution of 4K UHD (3,840 $\times$ 2,160). The incorporation of a high-spec screen ensures the optimal representation of the original image to the highest possible quality. During dataset collection, Tiny-ImageNet's subset images were displayed on the screen, using the left-top corner as the starting point.

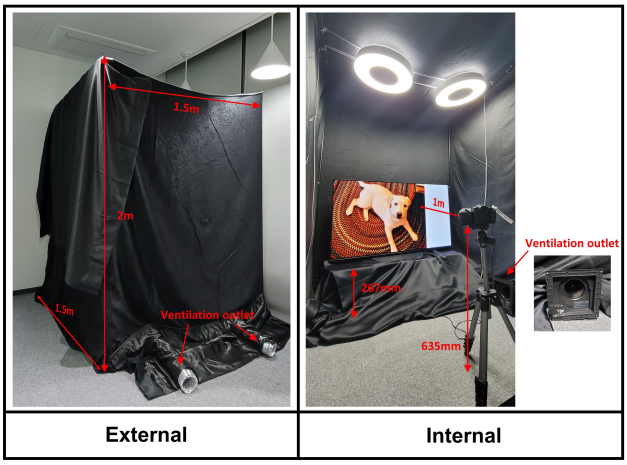


Figure 1. Actual appearance of *ES-Studio*

- **Ceiling Lamps (Component 2):** We have installed two 'Philips Hue White & Color Ambiance Infuse' lights, each with a maximum lumen[1] output of 3700 lm. We choose this model for its ability to provide sufficient brightness even in dark room, allowing for an appropriate depiction of a light-on scenario. Additionally, these ceiling lamps offer the advantage of automating dataset collection through remote control APIs. To prevent the issue of light reflecting on the screen, the screen and the camera are positioned at a sufficient distance from the ceiling lights.
- **Camera (Component 3):** The camera selected for ES-Studio is 'Canon EOS-RP' body paired with 'RF 24-105mm F4-7.1 IS STM' lens. When combining this lens and body configuration, ISO can be implemented in the range of 100 to 40000, shutter speed from 1/4000 to 30 seconds, and aperture from f4.0 to f22. We opted for a full-frame CMOS sensor model rather than a crop one to achieve a broader field of view and higher resolution. We acknowledge that a change in the camera, even with the same parameter settings (both manual and AE), can lead to variations in the captured image. In other words, a

[1] brightness emitted by a light source within a unit solid angle in one second

Table 1. Manual camera sensor parameter setting in validation set

Parameter No.	ISO	Shutter speed	Aperture
1	200	0”4’	f5.0
2	800	0”4’	f5.0
3	3200	0”4’	f5.0
4	12800	0”4’	f5.0
5	200	1/20’	f5.0
6	800	1/20’	f5.0
7	3200	1/20’	f5.0
8	12800	1/20’	f5.0
9	200	1/160’	f5.0
10	800	1/160’	f5.0
11	3200	1/160’	f5.0
12	12800	1/160’	f5.0
13	200	1/1250’	f5.0
14	800	1/1250’	f5.0
15	3200	1/1250’	f5.0
16	12800	1/1250’	f5.0
17	200	0”4’	f8.0
18	800	0”4’	f8.0
19	3200	0”4’	f8.0
20	12800	0”4’	f8.0
21	200	1/20’	f8.0
22	800	1/20’	f8.0
23	3200	1/20’	f8.0
24	12800	1/20’	f8.0
25	200	1/160’	f8.0
26	800	1/160’	f8.0
27	3200	1/160’	f8.0
28	12800	1/160’	f8.0
29	200	1/1250’	f8.0
30	800	1/1250’	f8.0
31	3200	1/1250’	f8.0
32	12800	1/1250’	f8.0
33	200	0”4’	f13
34	800	0”4’	f13
35	3200	0”4’	f13
36	12800	0”4’	f13
37	200	1/20’	f13
38	800	1/20’	f13
39	3200	1/20’	f13
40	12800	1/20’	f13
41	200	1/160’	f13
42	800	1/160’	f13
43	3200	1/160’	f13
44	12800	1/160’	f13
45	200	1/1250’	f13
46	800	1/1250’	f13
47	3200	1/1250’	f13
48	12800	1/1250’	f13
49	200	0”4’	f20
50	800	0”4’	f20
51	3200	0”4’	f20
52	12800	0”4’	f20
53	200	1/20’	f20
54	800	1/20’	f20
55	3200	1/20’	f20
56	12800	1/20’	f20
57	200	1/160’	f20
58	800	1/160’	f20
59	3200	1/160’	f20
60	12800	1/160’	f20
61	200	1/1250’	f20
62	800	1/1250’	f20
63	3200	1/1250’	f20
64	12800	1/1250’	f20

Table 2. Manual camera sensor parameter setting in test set

Parameter No.	ISO	Shutter speed	Aperture
1	250	1/4’	f5.0
2	2000	1/4’	f5.0
3	16000	1/4’	f5.0
4	250	1/60’	f5.0
5	2000	1/60’	f5.0
6	16000	1/60’	f5.0
7	250	1/1000’	f5.0
8	2000	1/1000’	f5.0
9	16000	1/1000’	f5.0
10	250	1/4’	f9.0
11	2000	1/4’	f9.0
12	16000	1/4’	f9.0
13	250	1/60’	f9.0
14	2000	1/60’	f9.0
15	16000	1/60’	f9.0
16	250	1/1000’	f9.0
17	2000	1/1000’	f9.0
18	16000	1/1000’	f9.0
19	250	1/4’	f16
20	2000	1/4’	f16
21	16000	1/4’	f16
22	250	1/60’	f16
23	2000	1/60’	f16
24	16000	1/60’	f16
25	250	1/1000’	f16
26	2000	1/1000’	f16
27	16000	1/1000’	f16

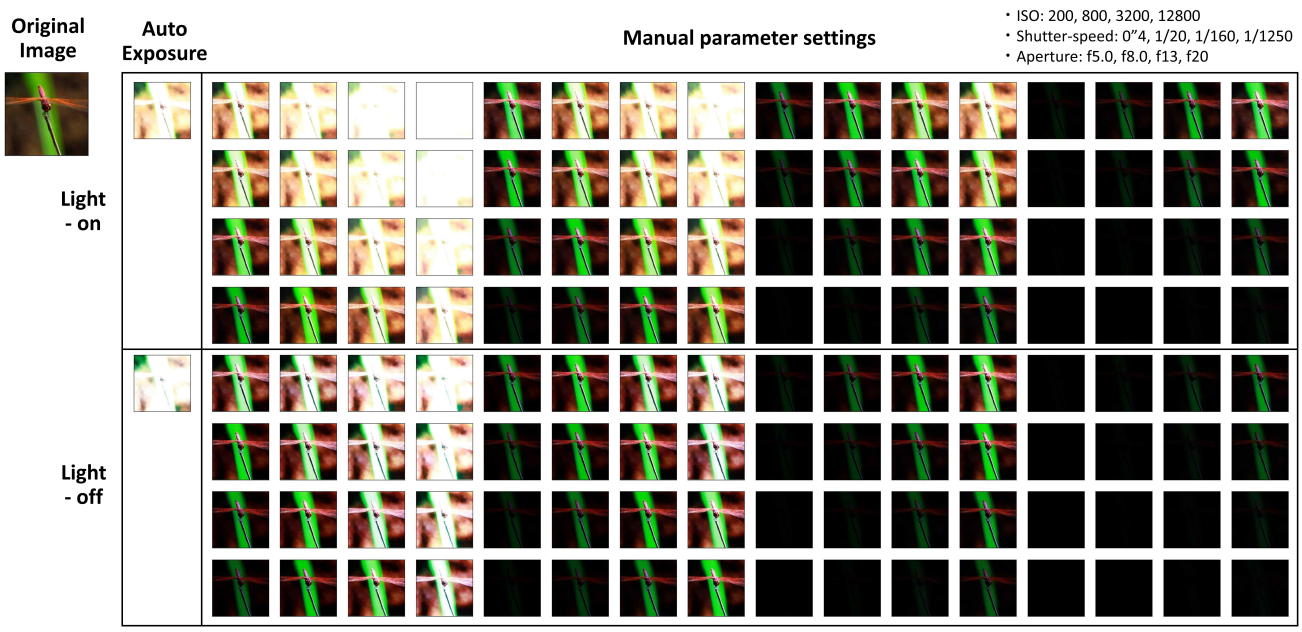
change in the camera’s hardware, even with identical software settings, can result in differences in the final output.

- **Desktop Computer (Component 4):** We automate the data collection system using the ‘Apple Mac Studio M2 Max’ desktop model, which communicates with the three aforementioned components via WIFI network. The desktop utilizes the Phillips Hue API for lighting control and the Canon camera control (CC) API for wireless camera control. The automation not only minimizes errors that could occur with human intervention, such as changes in camera position and external light interference, but also ensures consistency and accuracy, enabling faster and more efficient capturing and preprocessing processes.

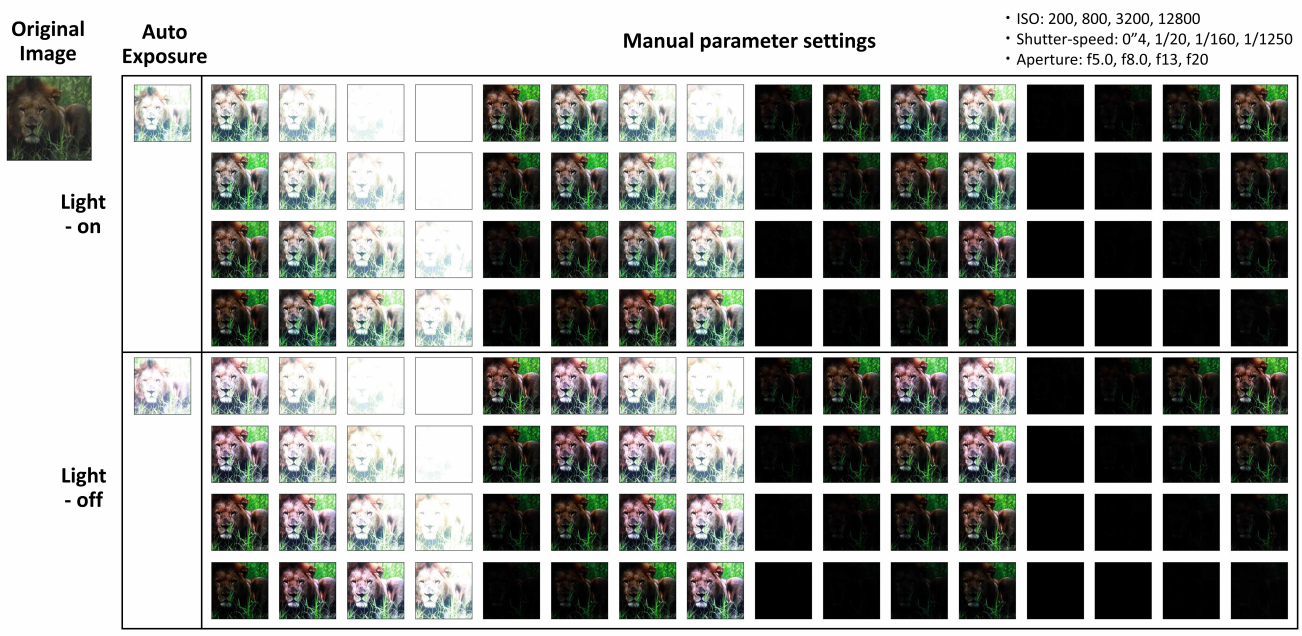
This comprehensive configuration ensures a controlled environment within *ES-Studio*, limiting external influences to only light factors and camera sensors.

1.2. Data Collection Module Implementation

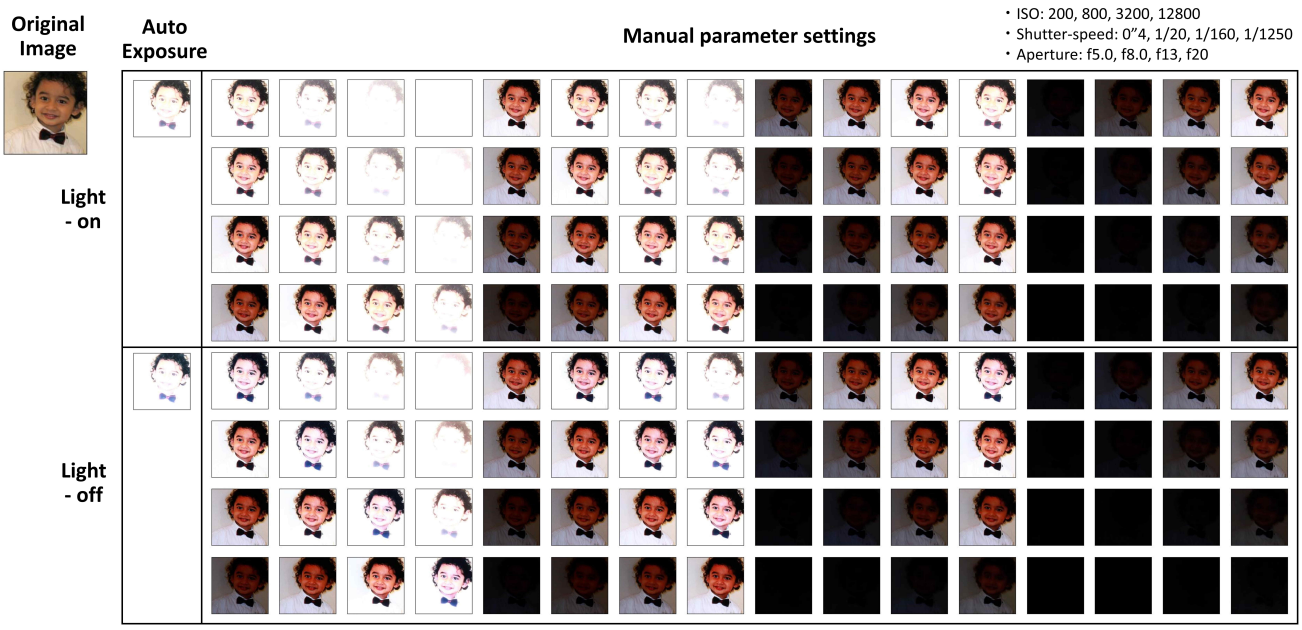
In this section, we revisit the key points discussed in Section 4.2. of the main text and subsequently delve into the finer details. In terms of reference dataset, a total of 2000 image samples were selected from a Tiny-ImageNet [24] subset. Out of these, 1000 images were dedicated to the validation set, and the remaining 1000 formed the test set, ensuring that there was no overlap between the two sets. During the data collection phase, each original reference image was


Original Image
Auto Exposure
Manual parameter settings
• ISO: 200, 800, 3200, 12800
• Shutter-speed: 0”4, 1/20, 1/160, 1/1250
• Aperture: f5.0, f8.0, f13, f20
Light - on
Light - off

(a)


Original Image
Auto Exposure
Manual parameter settings
• ISO: 200, 800, 3200, 12800
• Shutter-speed: 0”4, 1/20, 1/160, 1/1250
• Aperture: f5.0, f8.0, f13, f20
Light - on
Light - off

(b)



(c)

Figure 2. *ImageNet-ES* sample (a), (b), (c) : In all subfigures, a broad range of variations can be observed based on each parameter option. Looking at Figure 2a, the following becomes evident: under the same camera parameter settings (AE and manual), it is apparent that the captured image undergoes substantial changes depending on environmental variations (light on/off). Particularly in Figure 2c, variations in skin tone are noticeable due to changes in both environmental conditions (light on/off) and camera parameter settings. This indicates that alterations in camera sensor and environmental settings can bring about significant variations, especially in certain scenarios. For example, in contexts related to race, such changes may lead to semantic shifts.

taken with manual camera sensor parameter settings, spanning 64 options for the validation set and 27 options for the test set, as well as in auto exposure (AE) settings, the images were repeatedly captured five times, respectively. Notably, this procedure was reiterated for both the "light on" and "light off" environments, covering each of the (AE + 64) options for validation and (AE + 27) options for the test set in each lighting condition.

The detailed manual parameter settings are provided in the Tables 1 and 2. While determining manual parameter options, we aimed to evenly cover the ranges of each camera sensor parameter (*i.e.* ISO, shutter speed, aperture). However, scenarios involving shutter speeds exceeding 1 second were excluded, considering their infrequent occurrence in real-world situations. Additionally, the data collection process involved meticulous efforts to minimize distortion through precise camera angle adjustments and thorough attention to diverse camera settings. Regarding the focus, it has been set to AF (auto focus) mode, and the metering is set to evaluative metering mode, allowing the camera to assess the entire frame for metering[2] before determining the exposure. We have set the recording resolution during shooting to the maximum supported by the camera, which is approximately 26 million (6240×4160) pixels .

After undergoing the preprocessing steps described in Section 4.2.2 in the main paper, composite images of the 2000 reference images for the subjective validation process are generated. During the subjective validation, three reviewers individually assessed a set of 2000 composite images, documenting identified issues such as 1) crop errors, 2) missed images, and 3) label mismatches. In cases where even one reviewer detected minor issues, the appropriate measures and reevaluation were applied to the corresponding images. When deemed necessary, reshooting or reprocessing was carried out. This ensured the collection of accurate and consistent data. While visually inspecting the collected data, we were able to identify some interesting points. Those are on Figure 2.

2. Details on OOD Detection Experiments (Sec. 5.1)

2.1. Setup and Implementation

In this section, we provide more details regarding the experiments in Section 5.1. in the main paper. The datasets used for semantics-centric and Model-Specific OOD (MS-

[2]the process of how modern cameras decides to assign the right shutter speed and aperture based on the amount of light the camera can pick up

Table 3. Datasets used in OOD detection experiments

Experiment Setting	Train		Validation			Test			
	ID	OOD	ID	OOD	C-OOD	Near S-OOD	Far S-OOD	C-OOD	ID
Semantics-centric	S3	OpenImage-O (train)	S1	Textures (test)	$val_{ImageNet\text{-}ES}$ (128 options)	n/a	n/a	n/a	n/a
MS-OOD	S3+	S3−	S1+	S1−		SSB-hard [21], NINCO [2]	iNaturalist [20], Textures (test) [3], OpenImage-O [22]	$test_{ImageNet\text{-}ES}$ (54 options)	S2+

Table 4. Description of partitions of Tiny-ImageNet [24] validation set (10K samples)

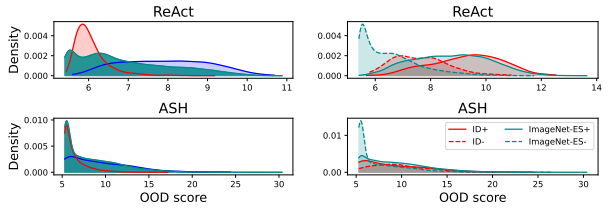
Partition	S1	S2	S3
	Reference of $val_{ImageNet\text{-}ES}$	Reference of $test_{ImageNet\text{-}ES}$	$(val_{Tiny-ImageNet} \setminus (S1 \cup S2))$
# of samples	1,000	1,000	8,000

Table 5. Description of underlying models for OOD detection experiments. (Optimizer: SGD, Scheduler: ReduceLROnPlateau, Batch size: 128)

Model	# of params	Pretrained	Acc. on Val_{Tin}	Training configuration
EfficientNet-B0 [19]	4.3M	ImageNet-1K	86.2%	lr: 5×10^{-3}, epochs: 20
ResNet18 [6]	11.3M		82.4%	lr: 5×10^{-2}, epochs: 15
ViT [5]	86M		91.2%	lr: 5×10^{-3} , epochs: 20
Swin-B [14]	86.9M		94.2%	lr: 5×10^{-3} , epochs: 20

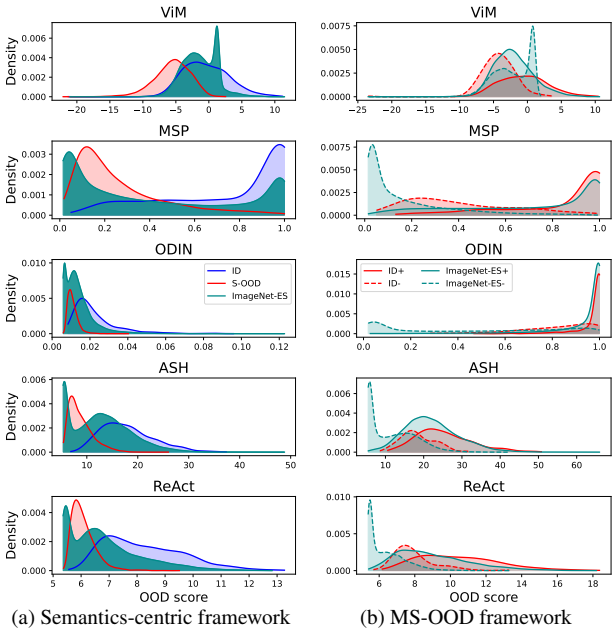
OOD) frameworks are outlined in Table 3. Other public datasets are used in its entirety, but we split the validation set of Tiny-ImageNet [24] into three segments: S1, S2, and S3. We pick the same images to the validation and test split of *ImageNet-ES* as S1 and S2, respectively. The remainder is designated as S3, which includes 40 images per each class. Since images in Tiny-ImageNet are provided in resized version (64 × 64), corresponding images from ImageNet are used to preserve the original resolution. This partition scheme of Tiny-ImageNet is described in Table 4. To train OOD detection methods within semantics-centric framework, we use S3 and the training set of OpenImage-O [22] as ID and OOD dataset, respectively. Within MS-OOD framework, S3+ and S3− are employed as ID and OOD datasets, respectively. To identify the validity of semantics-centric framework on *ImageNet-ES*, we employ S1 and the test set of Textures [3] as ID and OOD datasets respectivley. Within MS-OOD framework, we use S1+ and S1− as ID and OOD datasets, respectivley. For both framework, the validation set of *ImageNet-ES* is used as C-OOD dataset. We test five S-OODs (SSB-hard [21], NINCO [2], iNaturalist [20], Textures [3], OpenImage-O [22]) and categorize them into near-OOD and far-OOD following prior works [25]. We employed S2+ as ID in test set and assign samples in test set of *ImageNet-ES* into labels following each framework's policy and conducted tests.

The details of the underlying models are outlined in Table 5. This table also includes information of two additional models (Swin-B [14] and ResNet18 [6]) whose experimental results will be presented in the following section. All model weights used in the OOD detection experiments are obtained from timm library [23]. Since the obtained model weights produce prediction result for 1,000 classes, we finetune the classifier of each model to have 200 classes as in Tiny-ImageNet. Non-resized images from the training set



(a) Semantics-centric framework (b) MS-OOD framework

Figure 3. EfficientNet [19]: ReAct [18] and ASH [4] OOD score distribution with semantics-centric and MS-OOD frameworks. Tiny-ImageNet [24] and Texture [3] are used for the ID and S-OOD datasets, respectively. *ImageNet-ES* serves as a C-OOD dataset.

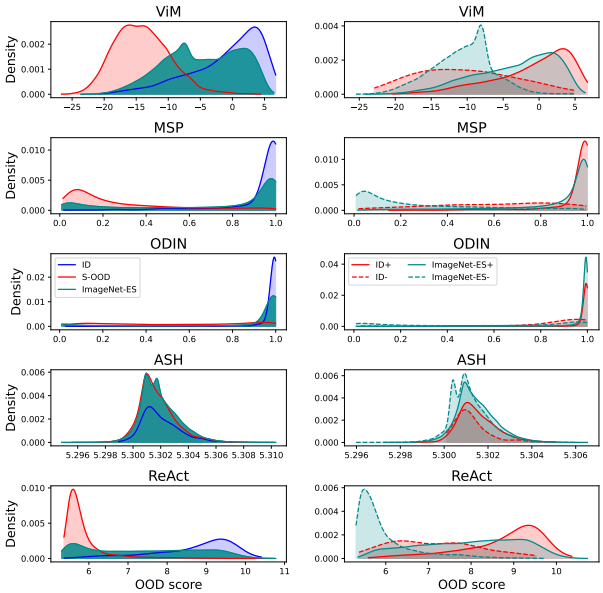


(a) Semantics-centric framework (b) MS-OOD framework

Figure 4. ResNet18 [6]: OOD score distribution with semantics-centric and MS-OOD frameworks. Tiny-ImageNet [24] and Texture [3] are used for the ID and S-OOD datasets, respectively. *ImageNet-ES* serves as a C-OOD dataset.

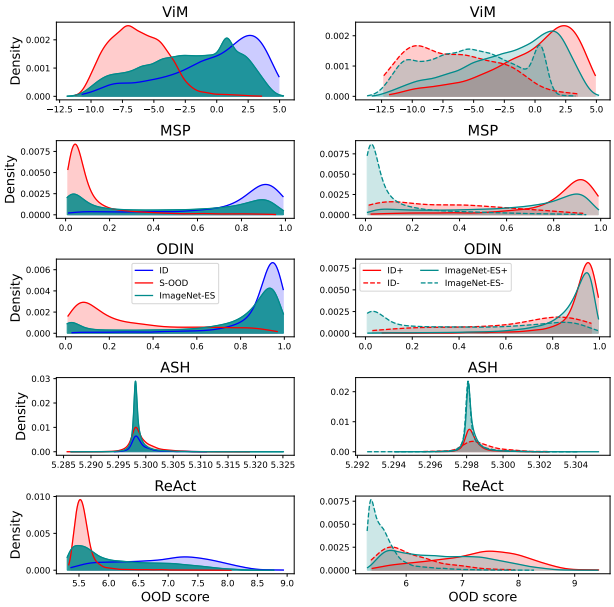
of Tiny-ImageNet are used for finetuning and the feature extractor part of each model is kept frozen during finetuning. The specific training configuration and final accuracy are also presented in Table 5.

To validate the current OOD detection methods, we fully



(a) Semantics-centric framework (b) MS-OOD framework

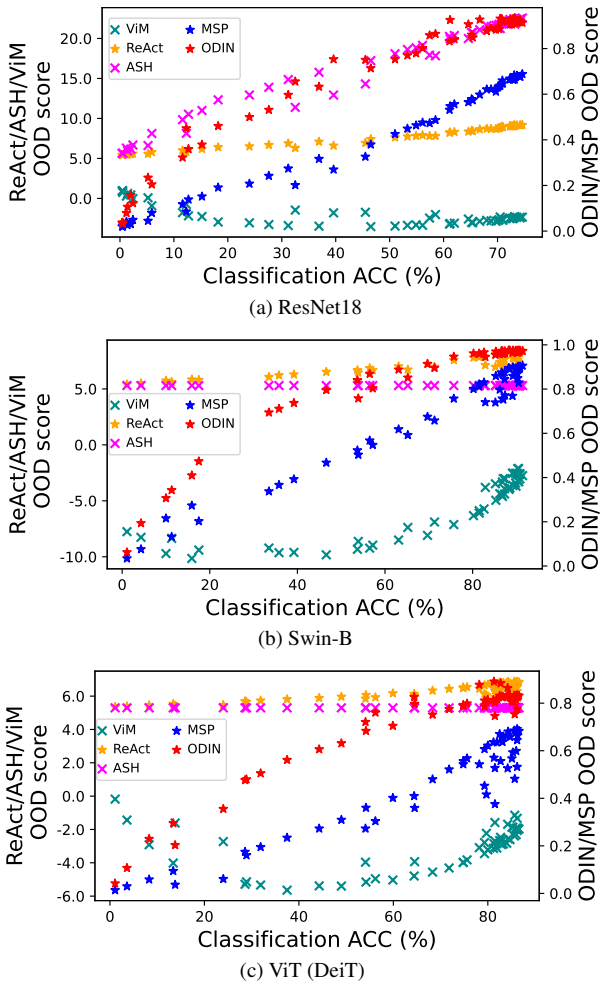
Figure 5. Swin-B [14]: OOD score distribution with semantics-centric and MS-OOD frameworks. Tiny-ImageNet [24] and Texture [3] are used for the ID and S-OOD datasets, respectively. *ImageNet-ES* serves as a C-OOD dataset.



(a) Semantics-centric framework (b) MS-OOD framework

Figure 6. ViT [5]: OOD score distribution with semantics-centric and MS-OOD frameworks. Tiny-ImageNet [24] and Texture [3] are used for the ID and S-OOD datasets, respectively. *ImageNet-ES* serves as a C-OOD dataset.

utilize the results and APIs provided by OpenOOD [25]. We experiment state-of-the-art methods, such as ViM [22] or ReAct [18] as well as MSP [8] which commonly serves as the baseline method. Advanced version of MSP, ODIN [13] is also tested. The implementation is based on



(a) ResNet18

(b) Swin-B

(c) ViT (DeiT)

Figure 7. Classification accuracy vs OOD score for additional underlying models. Each point represents the OOD score measured on the single parameter setting of *ImageNet-ES*

the OpenOOD package. All experiments are conducted on a system with Intel Xeon® Silver 4216 CPU and NVIDIA RTX 3090 GPU.

2.2. Additional Experiments

In this section, we present additional experimental findings for two additional models (ResNet18 [6], Swin-B [14], ViT [5]). The overall structure of experiments follows that presented in Section 5.1.1 and 5.1.2 of the main paper.

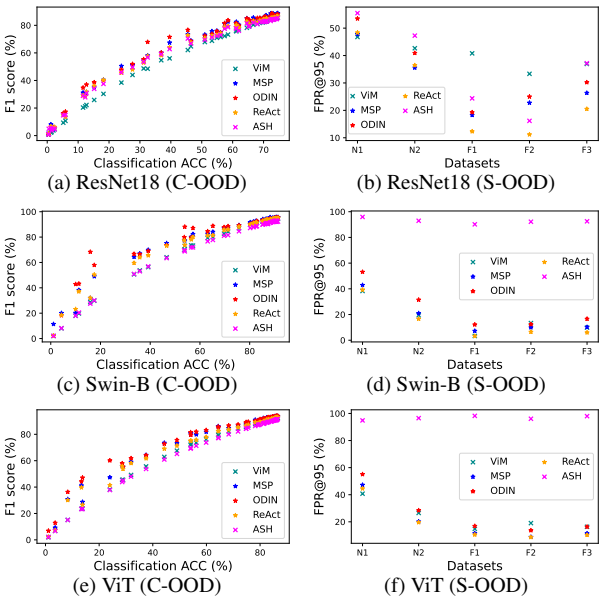
- **We verify that the MS-OOD framework [1] is still more suitable than the semantics-centric framework for other architectures as well.** Similar to the experiments done in Figure 5 of the main paper, we visualize the distribution of OOD score across different OOD detection methods within semantics-centric and MS-OOD framework. In Figure 4, 5 and 6 we observe consistent pattern in the OOD score distribution of ID, OOD and

ImageNet-ES (C-OOD) for ResNet18, Swin-B and ViT, as in EfficientNet. Furthermore, we evaluate 54 manual environmental/sensor variations in *ImageNet-ES* test set in terms of classification accuracy and OOD scores as in Figure 6 of the main paper, but based on different underlying models (ResNet18 [6], Swin-B [14] and ViT [5]). Figure 7 show both OOD score and the classification accuracy of different OOD detection methods in *ImageNet-ES*, based on Swin-B [14], ResNet18 [6] and and ViT [5], respectively. As in EfficientNet [19] case, we find that MSP [8] or ODIN [13] exhibit desirable correlation between OOD score and the classification accuracy, while ViM [22] demonstrates unacceptable results.

- We evaluate the performance of OOD detection methods based on different underlying models, similar to the experiments done in Figure 7 of the main paper. In Figure 8a, 8c and 8e, the C-OOD detection performance is presented in terms of F1 score. **We observe the tendency that closely resembles the pattern with EfficientNet [19] presented in the main paper**: MSP [8] and its advanced versions(ODIN [13] and ReAct [18]) consistently outperforms ASH [4] and ViM [22], which are the less effective than others.
- **Meanwhile, S-OOD performance exhibits different result when different underlying models are employed.** In Figure 8b, 8d and 8f we plot the S-OOD detection performance in terms of FPR. When Transformer-based models (Swin-B [14] and ViT [5]) are employed (Figure 8d, 8f), ASH [4] falls behind other methods across all S-OOD datasets. In the context of ResNet18 [6] (Figure 8b), ASH [4], demonstrate results improved than other models. In contrast, ViM [22] exhibits the performance notably inferior to the results obtained from EfficientNet.

3. Comparisions between MSP [8] and ViM [22]

MSP utilizes confidence scores and incurs OOD detection errors when the model overconfidently misclassifies a sample [18]. This event happens more frequently in S-OOD than in C-OOD since a misclassified C-OOD sample usually results in low confidence scores. On the other hand, ViM forms a feature space from ID samples and detects OOD when a sample contains a significant amount of unseen features [22]. However, while S-OOD samples are likely to contain unseen features, C-OOD samples present different distributions of already seen features. The lack of unseen features in C-OOD causes detection errors in ViM.



(a) ResNet18 (C-OOD) (b) ResNet18 (S-OOD)
(c) Swin-B (C-OOD) (d) Swin-B (S-OOD)
(e) ViT (C-OOD) (f) ViT (S-OOD)

Figure 8. Performance of OOD methods with C-OOD and S-OOD based on different underlying models. (a), (b) Each point is the F1 score measured on a parameter setting of $test_{ImageNet\text{-}ES}$. (c), (d) N1: SSB-hard [21], N2: NINCO [2], F1: iNaturalist [20], F2: Texture [3], F3: Openimage-O [22].

4. Details on Domain Generalization Experiments

4.1. Implementation Details of Finetuning

In this section, we provide the implementation details to produce the experimental results presented in Table 2 of the main paper. Every finetuning starts from the pretrained weights from PyTorch [17]. We use SGD optimizer with initial learning rate of 0.001. The learning rate decreases to 1e-6 following the cosine annealing schedule. We finetune for 10 epochs on each experiment. Batch size is set to 256.

During finetuning, we employ an image similarity metric LPIPS [26] to filter out some images that deviate too far from the original image. To do this, we calculate LPIPS between original image and perturbed image on each parameter setting. The calculated LPIPS is averaged over all 1,000 images collected under the same parameter setting. We use AlexNet [12] as a base model for LPIPS calculation. Calculated LPIPS for each manual parameter setting along with the sample image is presented in Figure 9. Based on the calculated LPIPS, we exclude images from parameter setting whose LPIPS is above the threshold which we set to 0.8. Chaning thershold would result in more/less perturbed images during finetuning. To study the impact, we experiment with different LPIPS threshold (0.6 and 1.0), as presented in Table 6. Utilizing less perturbed images (LPIPS threshold = 0.6) weakens the robustness of the model, on

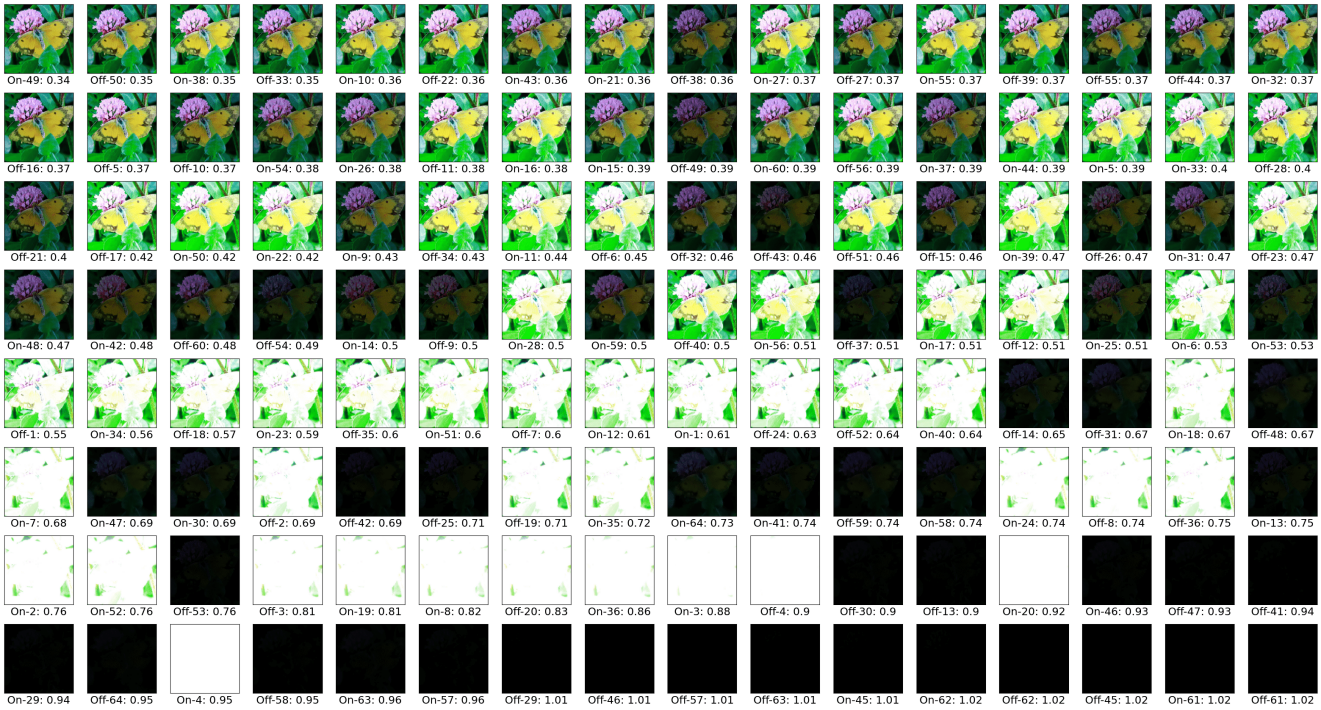


Figure 9. Calculated LPIPS [26] for each manual parameter setting with sample images on *ImageNet-ES*, in the ascending order of LPIPS. Environment, parameter setting number and LPIPS is provided under each image in the following format: [Environment]-[Parameter No.]: [LPIPS]

Table 6. Ablation study of finetuning techniques with different LPIPS threshold. The experiment ID and domain generalization techniques are same as those used in Table 2 of the main paper. The result is based on ResNet-50 [6]. (IN: ImageNet)

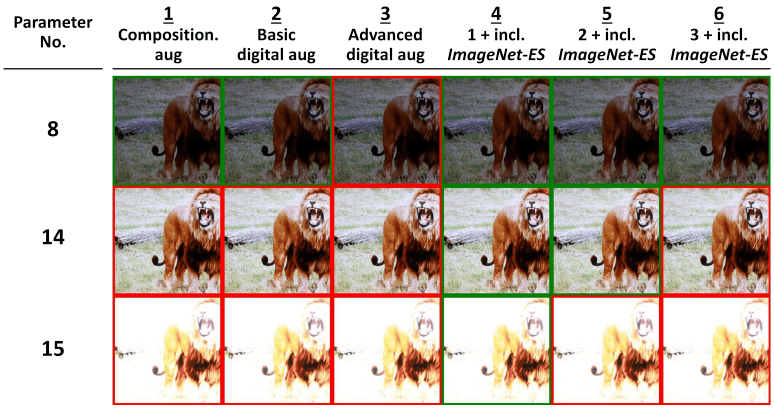
ID	Comp.aug	Basic digital aug	Advanced digital aug	Incl. *ImageNet-ES*	LPIPS threshold = 0.6			LPIPS threshold = 0.8			LPIPS threshold = 1.0		
					IN	IN-C	*ImageNet-ES*	IN	IN-C	*ImageNet-ES*	IN	IN-C	*ImageNet-ES*
4	✓			✓	85.8	51.4	53.9	86.0	51.8	55.8	85.8	52.0	56.3
5	✓	✓		✓	85.7	51.4	53.2	85.8	51.4	54.5	85.8	51.7	55.2
6	✓	✓	✓	✓	84.2	57.2	51.8	84.0	57.9	53.7	84.6	58.2	53.8

both digital (ImageNet-C [7]) and real world (*ImageNet-ES*) perturbations. Without digital augmentations, adding more perturbed images (LPIPS threshold = 1.0) results in degraded model performance on *ImageNet-ES* (Experiment 4). When finetuned with digital augmentations, however, the performance degradation could be diminished (Experiments 5 and 6).

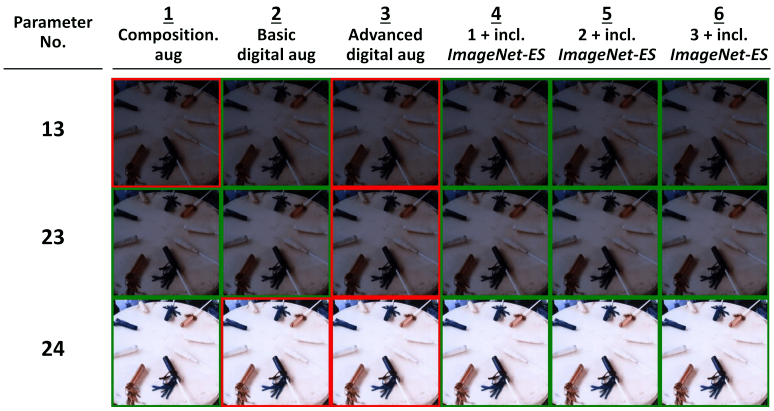
Composition-related and basic augmentations are implemented with functions provided by `torchvision`. Composition-related augmentations are implemented with `RandomResizedCrop` and `RandomHorizontalFlip`. Basic augmentations are implemented with `ColorJitter`, `RandomSolarize` and `RandomPosterize`. The implementation of DeepAugment [10] and AugMix [9] follows the open-sourced code from GitHub [10].

4.2. Qualitative Analysis of Domain Generalization Techniques

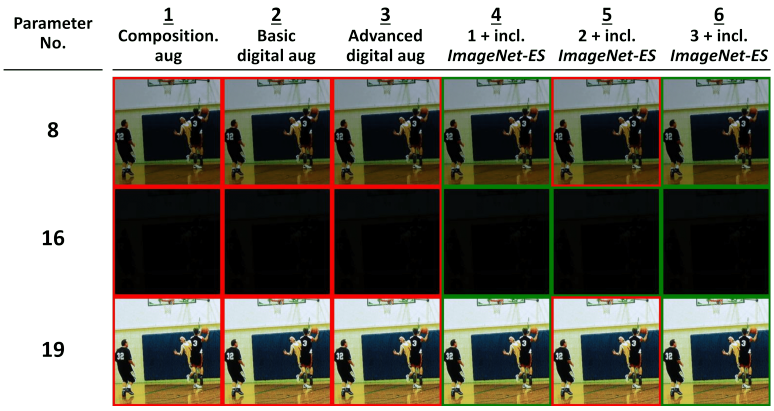
In Figure 10, we present the qualitative analysis of domain generalization techniques outlined in Table 2 in the main manuscript. Within each subfigure, we showcase the prediction result of each finetuned model for the selected example, each operating under chosen manual parameter settings. This analysis reveals that digital augmentation helps to improve the model performance on *ImageNet-ES* to some extent (parameter 13 in Figure 10b), but sometimes adversarially influences the model performance (parameter 8 in Figure 10a or parameter 24 in Figure 10b). Adding real-world perturbed images during finetuing proves beneficial in fostering model generalization across environmental and sensor domain as evidenced in Figure 10c. Furthermore, including images from *ImageNet-ES* helps to prevent the model degradation on environment and sensor domain (parameter 8 in Figure 10a or parameter 23, 24 in Figure 10b).



(a) Selected example 1 from *ImageNet-ES* (Class: n02129165)



(b) Selected example 2 from *ImageNet-ES* (Class: n03250847)



(c) Selected example 3 from *ImageNet-ES* (Class: n02802426)

Figure 10. Qualitative analysis of domain generalization techniques. Each subfigure depicts distinct parameter settings chosen from all manual parameters, when the light is on. The columns correspond to each finetuning strategy. The images correctly classified by the model are enclosed in green boxes, while those incorrectly classified are surrounded by red boxes.

5. Details on Evaluation Results of *ImageNet-ES*

In Tables 7-17, we provide more details of Table 3 in the main manuscript, the evaluation results of various models on *ImageNet-ES*: ResNet-50, DG version of ResNet-50, ResNet-152, EfficientNet-B0, EfficientNet-B3, SwinV2-T, SwinV2-B, OpenCLIP-b, OpenCLIP-h, DINOv2 (ViT-b) and DINOv2 (ViT-g). Each table provides the test accuracy of each model, evaluated at auto exposure and different manual parameter settings, under different environ-

Table 7. Detailed evaluation results of ResNet-50 [6] on *ImageNet-ES*.

Setting	Parameter No.	Environment		
		Light On	Light Off	Difference
Auto exposure	-	34.4	30.0	4.4
	Average	32.2		
Manual	1	49.1	49.3	0.2
	2	8.3	11.3	3.0
	3	0.6	1.6	1.0
	4	79.0	78.7	0.3
	5	62.3	66.7	4.4
	6	16.5	21.3	4.8
	7	57.9	50.5	7.4
	8	77.0	76.7	0.3
	9	74.0	75.0	1.0
	10	71.8	75.3	3.5
	11	26.4	31.0	4.6
	12	2.6	4.3	1.7
	13	78.1	76.7	1.4
	14	78.2	78.1	0.1
	15	40.8	46.4	5.6
	16	18.9	12.9	6.0
	17	69.3	63.8	5.5
	18	79.3	76.9	2.4
	19	79.3	80.1	0.8
	20	53.1	62.4	9.3
	21	9.5	14.1	4.6
	22	67.6	59.0	8.6
	23	79.3	78.5	0.8
	24	67.7	73.1	5.4
	25	1.6	0.7	0.9
	26	33.5	21.8	11.7
	27	73.5	68.8	4.7
	Best	79.3	80.1	0.1
		80.1		
	Worst	0.6	0.7	11.7
		0.6		
	Average	50.2	50.2	3.7
		50.2		

Table 8. Detailed evaluation results of DG version of ResNet-50 [6] on *ImageNet-ES*. This model is trained with DeepAugment [10] and AugMix [9] on ImageNet-21K

Setting	Parameter No.	Environment		
		Light On	Light Off	Difference
Auto exposure	-	55.6	51.0	4.5
	Average	53.3		
Manual	1	66.4	67.3	0.9
	2	22.5	25.8	3.3
	3	0.8	2.4	1.6
	4	83.6	83.3	0.3
	5	73.3	74.7	1.4
	6	36.7	40.2	3.5
	7	72.5	70.0	2.5
	8	82.3	81.8	0.5
	9	79.2	80.7	1.5
	10	79.2	80.2	1.0
	11	46.0	50.9	4.9
	12	5.7	11.3	5.6
	13	82.6	81.8	0.8
	14	82.3	82.6	0.3
	15	59.3	64.5	5.2
	16	54.7	45.5	9.2
	17	78.2	74.9	3.3
	18	82.6	81.8	0.8
	19	84.0	83.9	0.1
	20	68.5	72.4	3.9
	21	23.1	30.2	7.1
	22	77.6	73.2	4.4
	23	83.6	82.6	1.0
	24	74.4	79.1	4.7
	25	18.4	11.9	6.5
	26	60.5	50.2	10.3
	27	79.5	76.5	3.0
	Best	84.0	83.9	0.1
		84.0		
	Worst	0.8	2.4	10.3
		0.8		
	Average	61.4	61.5	3.2
		61.4		

ments (light on/off). The absolute value of difference between the test accuracy measured when light is on and off is also provided.

6. More analysis on camera parameters

As additional analysis, Figures 11 and 12 further motivate model-specific camera control. Compared to human-friendly (Worst and Auto) settings, the model-friendly (Best) setting provides different feature distributions (Figure 11) and more clearly clustered feature embeddings (Figure 12). Additionally, Figure 13 illustrates model performance according to solution candidates of camera control, revealing that the optimal parameters can vary with target models (Figure 13a) and samples (Figure 13b). Camera control can advance from aggregate-level to sample-wise immediate control.

Figure 11. Feature distributions: Model-friendly (Best) vs. Human-friendly (Worst, Auto)

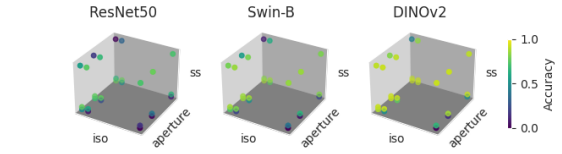
Figure 12. Feature embeddings: Model friendly (Best) vs. Human-friendly (Worst, Auto)

Table 9. Detailed evaluation results of ResNet-152 [6] on *ImageNet-ES*.

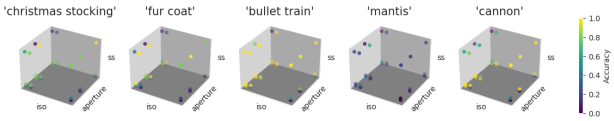
Setting	Parameter No.	Environment		
		Light On	Light Off	Difference
Auto exposure	-	42.6	39.6	3.0
	Average	41		
Manual	1	57.7	59.1	1.4
	2	10.3	14.8	4.5
	3	0.5	1.7	1.2
	4	82.6	82.7	0.1
	5	69.7	71.8	2.1
	6	22.7	26.1	3.4
	7	64.0	56.2	7.8
	8	81.2	80.5	0.7
	9	77.7	78.8	1.1
	10	76.3	78.8	2.5
	11	32.0	38.4	6.4
	12	3.1	6.1	3.0
	13	81.7	80.2	1.5
	14	80.5	81.5	1.0
	15	47.8	54.1	6.3
	16	23.0	15.2	7.8
	17	72.9	68.0	4.9
	18	81.3	80.3	1.0
	19	81.4	82.4	1.0
	20	61.1	69.1	8.0
	21	13.4	18.6	5.2
	22	73.1	64.6	8.5
	23	83.3	81.6	1.7
	24	72.8	77.2	4.4
	25	1.6	0.7	0.9
	26	39.7	24.8	14.9
	27	77.1	72.6	4.5
	Best	83.3	82.7	0.1
		83.3		
	Worst	0.5	0.7	14.9
		0.5		
	Average	54.4	54.3	3.9
		54.3		

Table 10. Detailed evaluation results of EfficientNet-B0 [19] on *ImageNet-ES*.

Setting	Parameter No.	Environment		
		Light On	Light Off	Difference
Auto exposure	-	53.8	49.0	4.8
	Average	51.4		
Manual	1	67.6	68.7	1.1
	2	19.7	22.9	3.2
	3	0.8	2.4	1.6
	4	83.6	83.0	0.6
	5	76.1	76.3	0.2
	6	35.8	39.2	3.4
	7	66.5	59.6	6.9
	8	82.4	81.7	0.7
	9	79.4	80.9	1.5
	10	79.0	80.6	1.6
	11	44.4	49.4	5.0
	12	5.1	10.5	5.4
	13	83.1	82.4	0.7
	14	82.4	82.8	0.4
	15	59.6	64.4	4.8
	16	21.9	15.1	6.8
	17	74.2	69.9	4.3
	18	81.6	80.9	0.7
	19	83.6	83.7	0.1
	20	70.3	74.5	4.2
	21	20.6	27.8	7.2
	22	75.0	68.0	7.0
	23	83.8	82.8	1.0
	24	77.1	79.0	1.9
	25	0.4	0.5	0.1
	26	39.4	24.6	14.8
	27	76.9	72.9	4.0
	Best	83.8	83.7	0.1
		83.8		
	Worst	0.4	0.5	14.8
		0.4		
	Average	58.2	57.9	3.3
		58.1		



(a) Model-wise solution space on 'Light-on'



(b) Sample-wise solution space on 'Light-on'

Figure 13. Solution space of camera parameter on *ImageNet-ES*

References

[1] Reza Averly and Wei-Lun Chao. Unified out-of-distribution detection: A model-specific perspective. *International Conference on Computer Vision (ICCV)*. 6

[2] Julian Bitterwolf, Maximilian Mueller, and Matthias Hein. In or out? fixing imagenet out-of-distribution detection evaluation. In *ICML*, 2023. 5, 7

[3] Mircea Cimpoi, Subhransu Maji, Iasonas Kokkinos, Sammy Mohamed, and Andrea Vedaldi. Describing textures in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014. 5, 6, 7

[4] Andrija Djurisic, Nebojsa Bozanic, Arjun Ashok, and Rosanne Liu. Extremely simple activation shaping for out-of-distribution detection. 2023. 5, 7

[5] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. *CoRR*, abs/2010.11929, 2020. 5, 6, 7

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 5, 6, 7, 8, 10, 11

[7] Dan Hendrycks and Thomas Dietterich. Benchmarking neural network robustness to common corruptions and perturbations. *International Conference on Learning Representa-*

Table 11. Detailed evaluation results of EfficientNet-B3 [19] on *ImageNet-ES*.

Setting	Parameter No.	Environment		
		Light On	Light Off	Difference
Auto exposure	-	64.2	59.7	4.5
	Average	62.0		
Manual	1	73.5	74.1	0.6
	2	27.6	33.8	6.2
	3	0.8	3.6	2.8
	4	85.8	86.2	0.4
	5	79.2	80.6	1.4
	6	45.6	48.8	3.2
	7	79.1	77.5	1.6
	8	84.7	84.0	0.7
	9	82.9	84.3	1.4
	10	82.0	82.5	0.5
	11	56.2	60.9	4.7
	12	8.6	14.5	5.9
	13	84.8	84.1	0.7
	14	84.9	85.6	0.7
	15	68.1	71.2	3.1
	16	64.2	54.9	9.3
	17	81.5	79.2	2.3
	18	84.0	83.7	0.3
	19	86.3	86.8	0.5
	20	74.2	78.1	3.9
	21	28.4	37.9	9.5
	22	82.2	80.5	1.7
	23	86.3	85.3	1.0
	24	80.3	82.4	2.1
	25	23.8	16.0	7.8
	26	68.9	56.5	12.4
	27	81.5	78.3	3.2
	Best	86.3	86.8	0.3
		86.8		
	Worst	0.8	3.6	12.4
		0.8		
	Average	66.1	66.3	3.3
		66.2		

Table 12. Detailed evaluation results of SwinV2-T [15] on *ImageNet-ES*.

Setting	Parameter No.	Environment		
		Light On	Light Off	Difference
Auto exposure	-	56.8	51.7	5.1
	Average	54.2		
Manual	1	69.6	70.8	1.2
	2	20.0	23.9	3.9
	3	0.6	2.2	1.6
	4	86.4	86.1	0.3
	5	77.9	79.9	2.0
	6	34.8	39.4	4.6
	7	78.5	74.7	3.8
	8	86.1	84.6	1.5
	9	82.8	84.1	1.3
	10	83.2	84.9	1.7
	11	45.4	49.6	4.2
	12	5.7	9.3	3.6
	13	85.8	85.4	0.4
	14	85.2	86.2	1.0
	15	61.3	65.9	4.6
	16	55.5	45.3	10.2
	17	80.7	77.4	3.3
	18	84.9	83.2	1.7
	19	86.8	86.2	0.6
	20	72.4	76.0	3.6
	21	20.2	28.6	8.4
	22	83.0	80.5	2.5
	23	86.0	85.8	0.2
	24	79.3	83.6	4.3
	25	13.5	8.4	5.1
	26	60.7	45.5	15.2
	27	80.1	76.0	4.1
	Best	86.8	86.2	0.2
		86.8		
	Worst	0.6	2.2	15.2
		0.6		
	Average	63.2	63.1	3.5
		63.1		

tions, 2019. 8

[8] Dan Hendrycks and Kevin Gimpel. A baseline for detecting misclassified and out-of-distribution examples in neural networks. 2017. 6, 7

[9] Dan Hendrycks, Norman Mu, Ekin D. Cubuk, Barret Zoph, Justin Gilmer, and Balaji Lakshminarayanan. AugMix: A simple data processing method to improve robustness and uncertainty. *Proceedings of the International Conference on Learning Representations (ICLR)*, 2020. 8, 10

[10] Dan Hendrycks, Steven Basart, Norman Mu, Saurav Kadavath, Frank Wang, Evan Dorundo, Rahul Desai, Tyler Zhu, Samyak Parajuli, Mike Guo, et al. The many faces of robustness: A critical analysis of out-of-distribution generalization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8340–8349, 2021. 8, 10

[11] Gabriel Ilharco, Mitchell Wortsman, Ross Wightman, Cade Gordon, Nicholas Carlini, Rohan Taori, Achal Dave, Vaishaal Shankar, Hongseok Namkoong, John Miller, Hannaneh Hajishirzi, Ali Farhadi, and Ludwig Schmidt. Openclip, 2021. If you use this software, please cite it as below. 13, 14

[12] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012. 7

[13] Shiyu Liang, Yixuan Li, and R. Srikant. Enhancing the reliability of out-of-distribution image detection in neural networks. 2018. 6, 7

[14] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 10012–10022, 2021. 5, 6, 7

[15] Ze Liu, Han Hu, Yutong Lin, Zhuliang Yao, Zhenda Xie, Yixuan Wei, Jia Ning, Yue Cao, Zheng Zhang, Li Dong, et al. Swin transformer v2: Scaling up capacity and resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12009–12019, 2022. 12, 13

[16] Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy

Table 13. Detailed evaluation results of SwinV2-B [15] on *ImageNet-ES*.

Setting	Parameter No.	Environment		
		Light On	Light Off	Difference
Auto exposure	-	62.3	58.0	4.3
	Average	60.1		
Manual	1	75.2	76.3	1.1
	2	22.6	27.3	4.7
	3	0.6	2.1	1.5
	4	88.3	87.8	0.5
	5	82.5	83.5	1.0
	6	39.8	44.0	4.2
	7	81.3	77.2	4.1
	8	86.6	85.4	1.2
	9	85.7	86.8	1.1
	10	86.6	86.9	0.3
	11	51.2	57.3	6.1
	12	6.3	11.3	5.0
	13	87.0	87.2	0.2
	14	88.2	88.0	0.2
	15	67.2	72.0	4.8
	16	57.4	44.5	12.9
	17	82.3	80.0	2.3
	18	86.6	84.6	2.0
	19	89.0	87.6	1.4
	20	76.9	80.4	3.5
	21	23.1	32.7	9.6
	22	84.1	81.9	2.2
	23	87.3	86.2	1.1
	24	83.5	85.7	2.2
	25	9.7	6.0	3.7
	26	62.2	46.5	15.7
	27	82.3	80.7	1.6
	Best	89.0	88.0	0.2
		89.0		
	Worst	0.6	2.1	15.7
		0.6		
	Average	65.7	65.6	3.5
		65.6		

Table 14. Detailed evaluation results of OpenCLIP-b [11] on *ImageNet-ES*.

Setting	Parameter No.	Environment		
		Light On	Light Off	Difference
Auto exposure	-	68.5	64.1	4.4
	Average	66.3		
Manual	1	80.8	82.4	1.6
	2	30.2	34.7	4.5
	3	1.1	3.5	2.4
	4	92.7	92.5	0.2
	5	87.4	88.0	0.6
	6	47.2	50.9	3.7
	7	85.8	83.3	2.5
	8	91.8	91.2	0.6
	9	90.4	91.2	0.8
	10	90.1	90.9	0.8
	11	58.2	63.5	5.3
	12	7.1	14.0	6.9
	13	92.3	91.5	0.8
	14	91.5	92.1	0.6
	15	73.1	78.1	5.0
	16	65.8	54.4	11.4
	17	89.5	85.6	3.9
	18	91.2	90.6	0.6
	19	92.4	92.4	0.0
	20	82.3	85.1	2.8
	21	29.8	40.3	10.5
	22	89.7	86.6	3.1
	23	91.8	91.3	0.5
	24	88.9	90.0	1.1
	25	14.8	8.2	6.6
	26	71.4	58.1	13.3
	27	88.9	87.0	1.9
	Best	92.7	92.5	0.0
		92.7		
	Worst	1.1	3.5	13.3
		1.1		
	Average	71.0	71.0	3.4
		71.0		

Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*, 2023. 14, 15

[17] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019. 7

[18] Yiyou Sun, Chuan Guo, and Yixuan Li. React: Out-of-distribution detection with rectified activations. *Advances in Neural Information Processing Systems*, 34:144–157, 2021. 5, 6, 7

[19] Mingxing Tan and Quoc Le. EfficientNet: Rethinking model scaling for convolutional neural networks. In *Proceedings of the 36th International Conference on Machine Learning*, pages 6105–6114. PMLR, 2019. 5, 7, 11, 12

[20] Grant Van Horn, Oisin Mac Aodha, Yang Song, Yin Cui, Chen Sun, Alex Shepard, Hartwig Adam, Pietro Perona, and Serge Belongie. The inaturalist species classification and detection dataset. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 5, 7

[21] Sagar Vaze, Kai Han, Andrea Vedaldi, and Andrew Zisserman. Open-set recognition: A good closed-set classifier is all you need. In *International Conference on Learning Representations*, 2022. 5, 7

[22] Haoqi Wang, Zhizhong Li, Litong Feng, and Wayne Zhang. Vim: Out-of-distribution with virtual-logit matching supplementary material. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4921–4930, 2022. 5, 6, 7

[23] Ross Wightman. Pytorch image models. `https://github.com/rwightman/pytorch-image-models`, 2019. 5

[24] Jiayu Wu, Qixiang Zhang, and Guoxi Xu. Tiny imagenet challenge. *Technical report*, 2017. 2, 5, 6

[25] Jingyang Zhang, Jingkang Yang, Pengyun Wang, Haoqi Wang, Yueqian Lin, Haoran Zhang, Yiyou Sun, Xuefeng Du,

Table 15. Detailed evaluation results of OpenCLIP-h [11] on *ImageNet-ES*.

Setting	Parameter No.	Environment		
		Light On	Light Off	Difference
Auto exposure	-	80.2	78.0	2.2
	Average	79.1		
Manual	1	86.9	87.2	0.3
	2	39.1	44.7	5.6
	3	1.1	5.2	4.1
	4	93.6	93.8	0.2
	5	92.4	92.1	0.3
	6	60.4	64.2	3.8
	7	92.1	90.2	1.9
	8	93.7	93.3	0.4
	9	93.4	93.2	0.2
	10	92.6	93.7	1.1
	11	71.5	75.2	3.7
	12	12.6	19.0	6.4
	13	93.8	93.7	0.1
	14	93.7	93.5	0.2
	15	83.1	86.4	3.3
	16	80.9	74.7	6.2
	17	93.2	91.2	2.0
	18	93.5	93.6	0.1
	19	93.9	94.7	0.8
	20	88.6	91.3	2.7
	21	39.7	52.7	13.0
	22	93.4	92.2	1.2
	23	93.8	93.5	0.3
	24	92.1	93.5	1.4
	25	43.0	29.8	13.2
	26	84.3	76.3	8.0
	27	92.8	90.9	1.9
	Best	93.9	94.7	0.1
		94.7		
	Worst	1.1	5.2	13.2
		1.1		
	Average	77.4	77.8	3.1
		77.6		

Kaiyang Zhou, Wayne Zhang, et al. Openood v1. 5: Enhanced benchmark for out-of-distribution detection. *arXiv preprint arXiv:2306.09301*, 2023. 5, 6

[26] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018. 7, 8

Table 16. Detailed evaluation results of DINOv2 (ViT-b) [16] on *ImageNet-ES*.

Setting	Parameter No.	Environment		
		Light On	Light Off	Difference
Auto exposure	-	76.0	72.9	3.1
	Average	74.5		
Manual	1	85.1	85.3	0.2
	2	34.6	38.9	4.3
	3	1.2	4.6	3.4
	4	92.2	91.9	0.3
	5	89.8	90.4	0.6
	6	53.1	58.6	5.5
	7	88.0	86.3	1.7
	8	91.3	91.5	0.2
	9	91.5	91.8	0.3
	10	91.6	91.7	0.1
	11	66.7	71.7	5.0
	12	11.1	17.2	6.1
	13	91.7	91.8	0.1
	14	92.1	91.7	0.4
	15	80.3	84.5	4.2
	16	74.1	64.6	9.5
	17	89.6	87.7	1.9
	18	91.6	90.9	0.7
	19	91.6	91.8	0.2
	20	86.0	88.0	2.0
	21	35.4	45.6	10.2
	22	90.2	88.1	2.1
	23	91.4	91.2	0.2
	24	90.3	90.7	0.4
	25	25.3	14.6	10.7
	26	78.6	67.0	11.6
	27	89.1	86.9	2.2
	Best	92.2	91.9	0.1
		92.2		
	Worst	1.2	4.6	11.6
		1.2		
	Average	73.8	73.9	3.1
		73.9		

Table 17. Detailed evaluation results of DINOv2 (ViT-g) [16] on *ImageNet-ES*.

Setting	Parameter No.	Environment		
		Light On	Light Off	Difference
Auto exposure	-	85.5	83.1	2.4
	Average	84.3		
Manual	1	90.4	91.0	0.6
	2	49.8	52.8	3.0
	3	1.3	7.0	5.7
	4	93.7	93.6	0.1
	5	94.0	93.8	0.2
	6	68.5	72.6	4.1
	7	92.2	90.5	1.7
	8	93.6	93.3	0.3
	9	93.6	93.8	0.2
	10	93.5	94.2	0.7
	11	77.8	83.9	6.1
	12	15.0	25.1	10.1
	13	94.0	93.8	0.2
	14	93.8	94.1	0.3
	15	88.5	90.3	1.8
	16	83.0	76.6	6.4
	17	92.2	91.1	1.1
	18	93.7	93.1	0.6
	19	94.0	94.2	0.2
	20	92.2	93.1	0.9
	21	48.8	61.9	13.1
	22	92.4	91.7	0.7
	23	93.9	93.8	0.1
	24	93.9	94.0	0.1
	25	44.6	26.4	18.2
	26	84.7	78.9	5.8
	27	92.1	90.9	1.2
	Best	94.0	94.2	0.1
		94.2		
	Worst	1.3	7.0	18.2
		1.3		
	Average	79.5	79.8	3.1
		79.6		