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

CVPR SEATTLE, WA JUNE 17-21, 2024

2024-06-19

Eunsu Baek, Keondo Park, Jiyoon Kim, Hyung-sin Kim

Seoul National University
Graduate School of Data Science

Motivation: Domain shift

- DNN models are widely used recently.
- They, however, often does not perform well when **environment changes!**
- **Domain shift** is the main reason for the performance drop.
  - *Training data domain ≠ Test data domain*

Video shows 8-car pileup after a Tesla allegedly using Full Self-Driving stopped in a highway tunnel

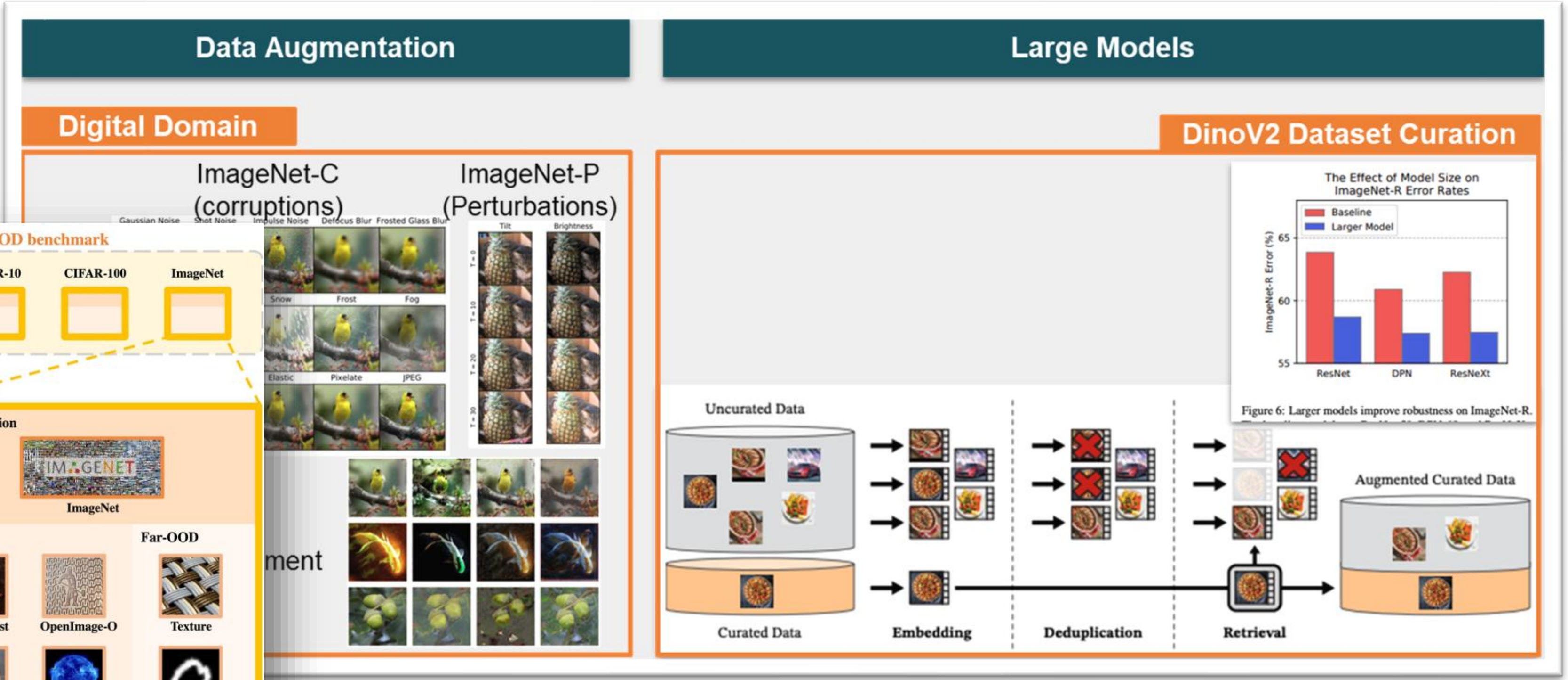
Grace Kay Jan 11, 2023, 8:00 AM GMT+9 Share Save

Video obtained by The Intercept appears to show a Tesla causing an 8-car pileup on November 24, 2022. The Intercept reporter Ken Klippenstein on Twitter

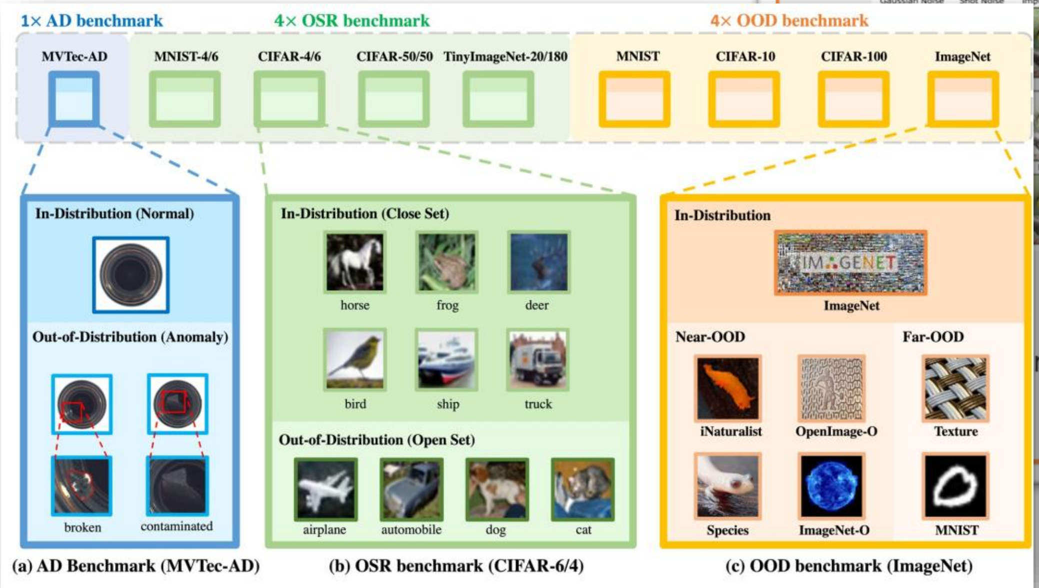
Source: Business insider

Motivation: Domain shift

- Many approaches to tackle domain shift.



Domain adaptation / Generalziation

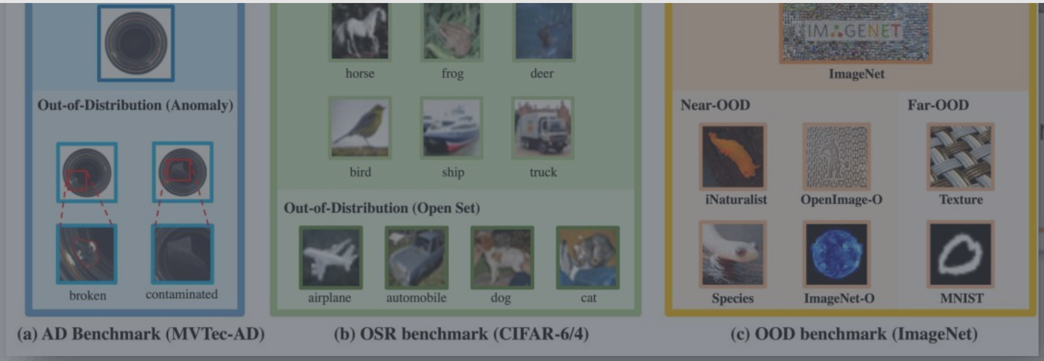


Out-of-Distribution (OOD) detection

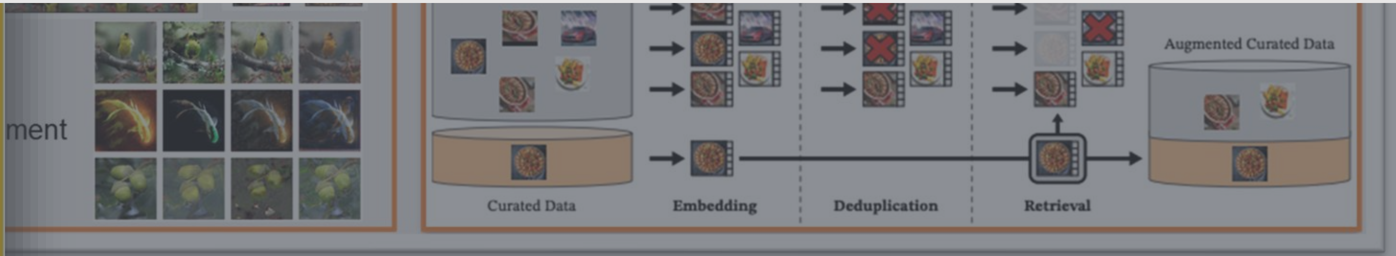
Motivation: Domain shift

- Many approaches to tackle domain shift.

Tries to make a smarter DNN model, but is it the best?



Out-of-Distribution (OOD) detection



Domain adaptation / Generalziation

Motivation: Domain shift

- Human vision system

Motivation: Domain shift

- Human vision system

Current solution

Read more books and be smart :)

It's too bright!!
I can't see anything!!

It's too close!!
I can't see anything!!

Motivation: Domain shift

- Human vision system

Current

“Best solution?"

It’s too bright!!
I can’t see anything!!

It’s too close!!
I can’t see anything!!

Motivation: Domain shift

- Human vision system

Current

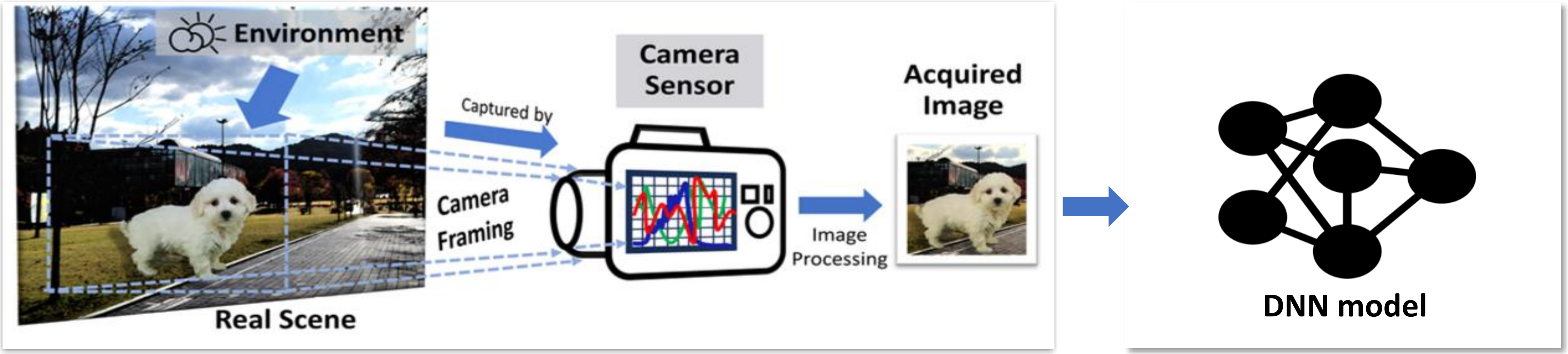
"Easier and better solution!"

or

Motivation: Domain shift

- Computer vision system

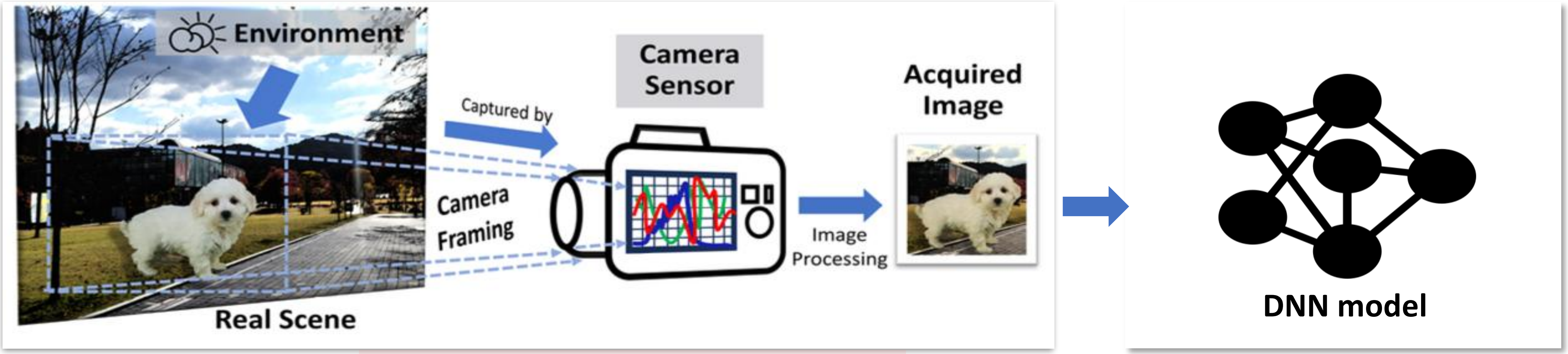
Real-world image acquisition process	Scene understanding



Motivation: Domain shift

- Computer vision system

Real-world image acquisition process | **Scene understanding**



	Composition Domain	Environment / Sensor Domain		Digital Domain	
Perturbation	ImageNet -R, A, E	***ImageNet-ES* (ours)**		ImageNet -C, P	Deep-Augment
Prediction Pipeline	**Real Scene + Camera Framing** →	**Environment** (Light) →	**Camera Sensor** (ISO, Aperture, Shutter-speed) →	**Acquired Image** →	**Model** → **Prediction**

ES-Studio

- Controllable testbed for **E**nvironment and **S**ensor domain
  - Capture real-world perturbations related to **light** (On/Off) and **camera parameters** (ISO / Shutter speed / Aperture)
  - Ensure reproducibility

[ES-Studio Design Description]

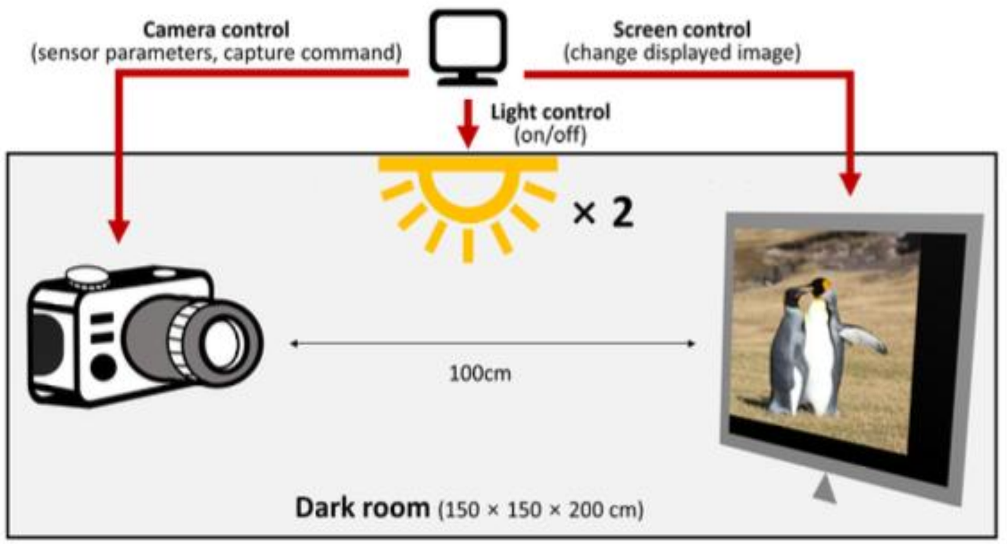
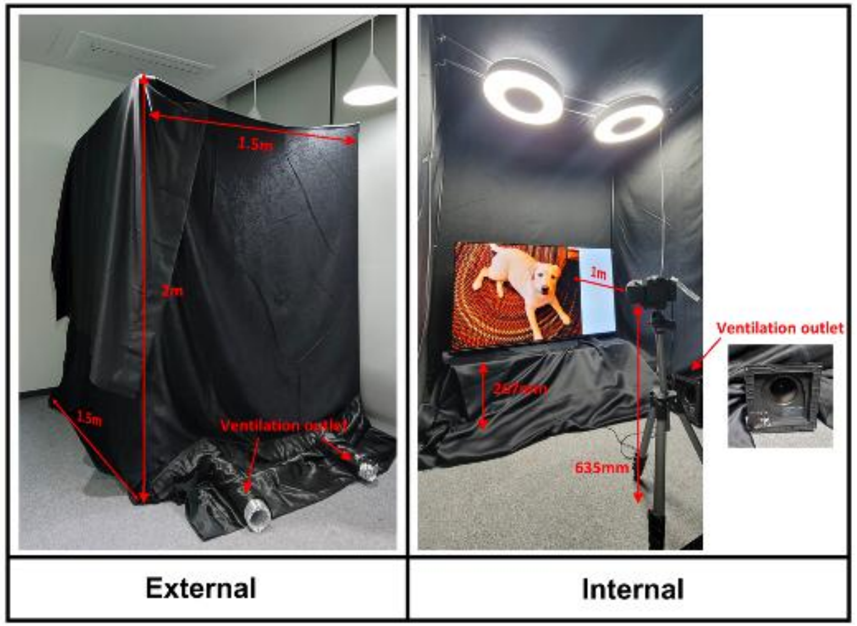


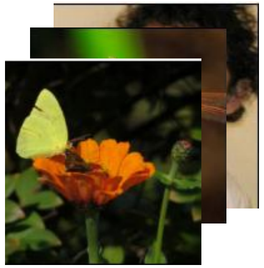
Figure 4. Illustration of the *ES-Studio* setup

[Photo of ES-Studio]



ImageNet-ES dataset

- Covariate shift datasets from the environment & sensor domain



ISO | Shutter Speed | Aperture

Auto Exposure (5 shots)

Val. Set 1000 sampled images from ImageNet ✕ 2 Light options (On / Off) ✕ (64 + 5) Camera parameter options = 138K Images

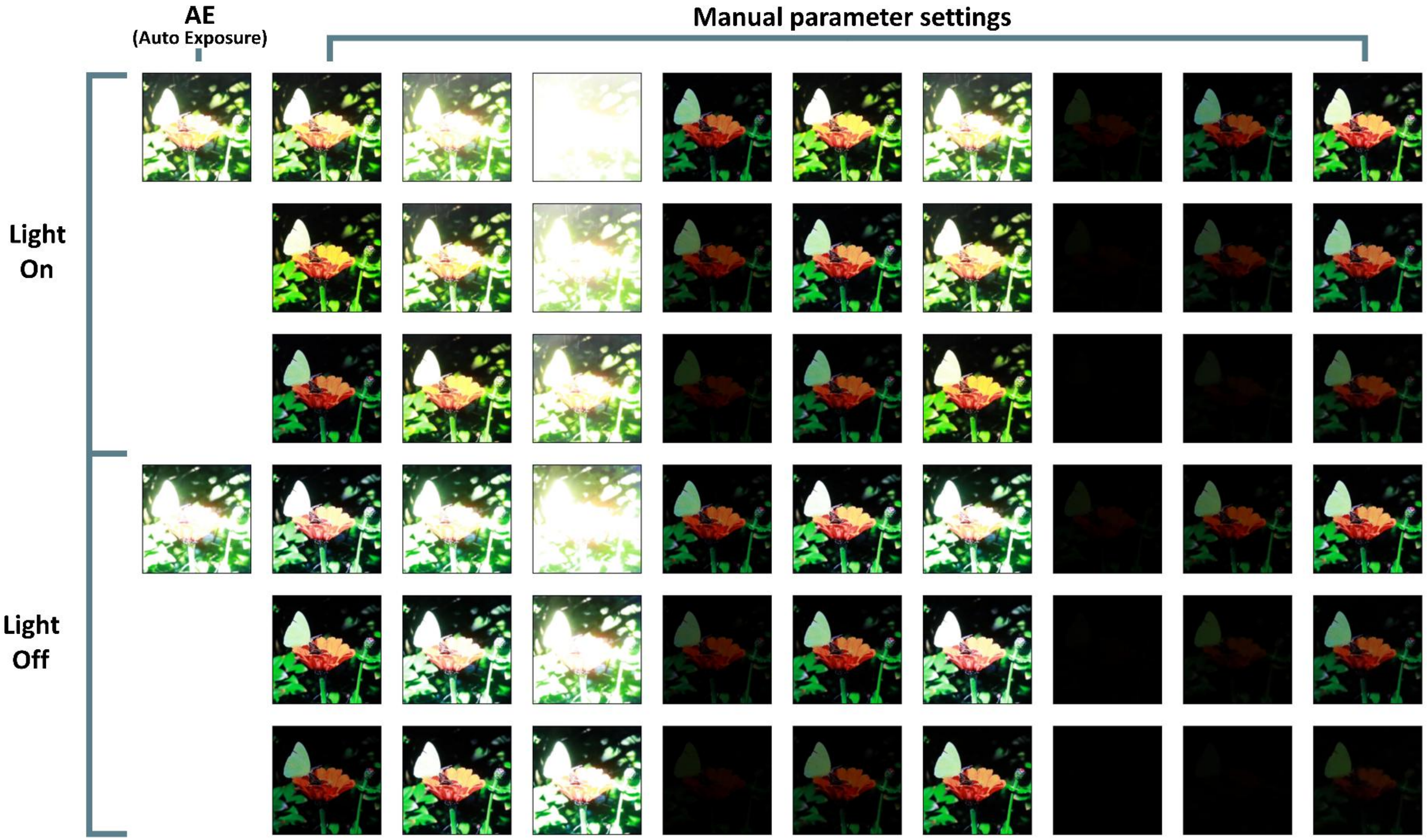
Test Set 1000 sampled images from ImageNet ✕ 2 Light options (On / Off) ✕ (27 + 5) Camera parameter options = 64K Images

Total 202K Images

ImageNet-ES dataset

- Sample images from test set:

Original Image
(From ImageNet)

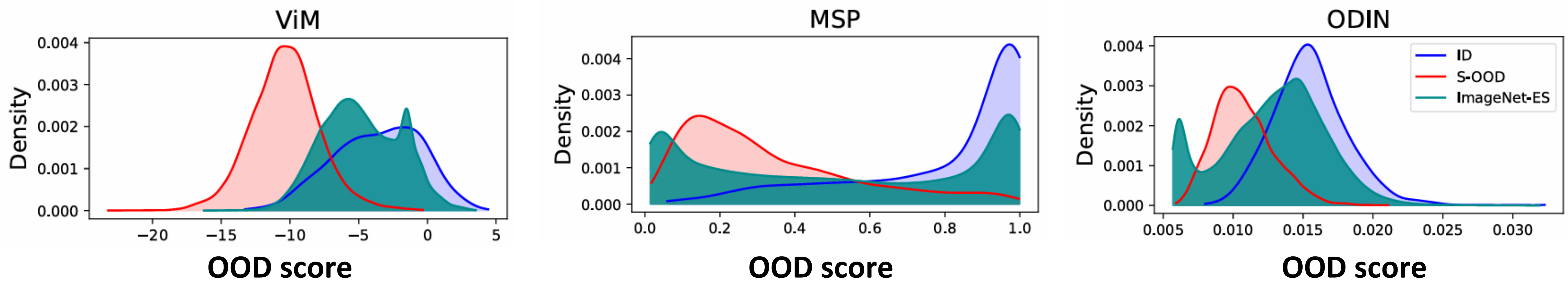


Experiments: Out-of-Distribution (OOD) detection

- What is the best OOD definition?
- Semantics-centric framework
  - Most widely used.
  - Any samples not included in the class definition of training domain => OOD
  - Treating C-OOD (Covariate shifted data, e.g. *ImageNet-ES*) as OOD or ID in entirety.
  - SOTA OOD detection techniques (ViM, ODIN, etc.) developed to work well under this framework.

Experiments: Out-of-Distribution (OOD) detection

- Semantics-centric framework
  - Test on three OOD detection techniques: ViM, MSP, ODIN
  - Model: EfficientNet-B0
  - Three datasets
    - In-Distribution (ID): Tiny-ImageNet
    - Semantics OOD (S-OOD): Texture-O
    - Covariate shifted OOD (C-OOD): *ImageNet-ES*



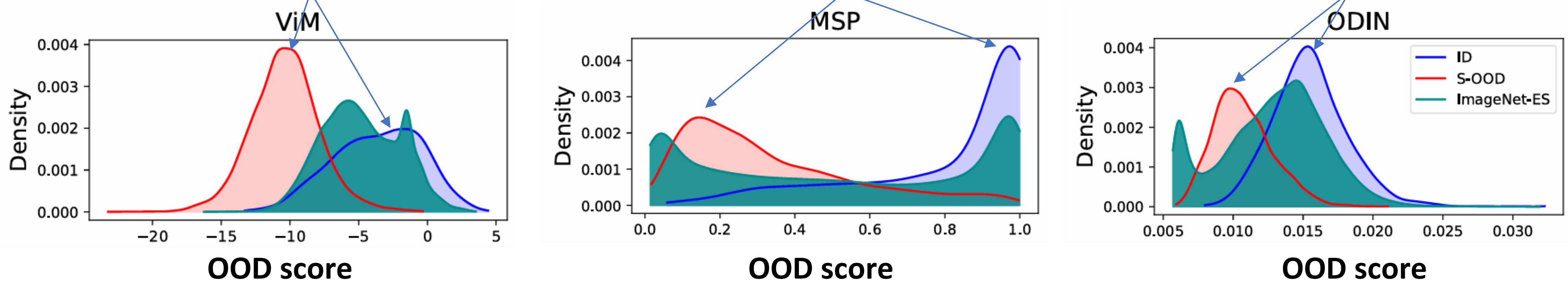
Experiments: Out-of-Distribution (OOD) detection

- Semantics-centric framework
  - Test on three OOD detection techniques: ViM, MSP, ODIN
  - Model: EfficientNet-B0
  - Three datasets
    - In-Distribution (ID): Tiny-ImageNet
    - Semantics OOD (S-OOD): Texture-O
    - Covariate shifted OOD (C-OOD): *ImageNet-ES*

Clear separation between ID and S-OOD

Clear separation between ID and S-OOD

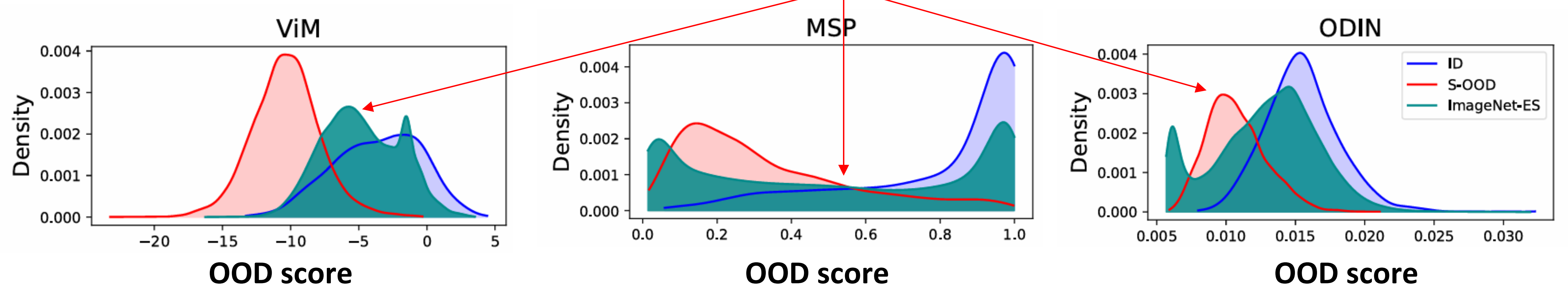
Clear separation between ID and S-OOD



Experiments: Out-of-Distribution (OOD) detection

- Semantics-centric framework
  - Test on three OOD detection techniques: ViM, MSP, ODIN
  - Model: EfficientNet-B0
  - Three datasets
    - In-Distribution (ID): Tiny-ImageNet
    - Semantics OOD (S-OOD): Texture-O
    - Covariate shifted OOD (C-OOD): *ImageNet-ES*

No clear distinction on ImageNet-ES!

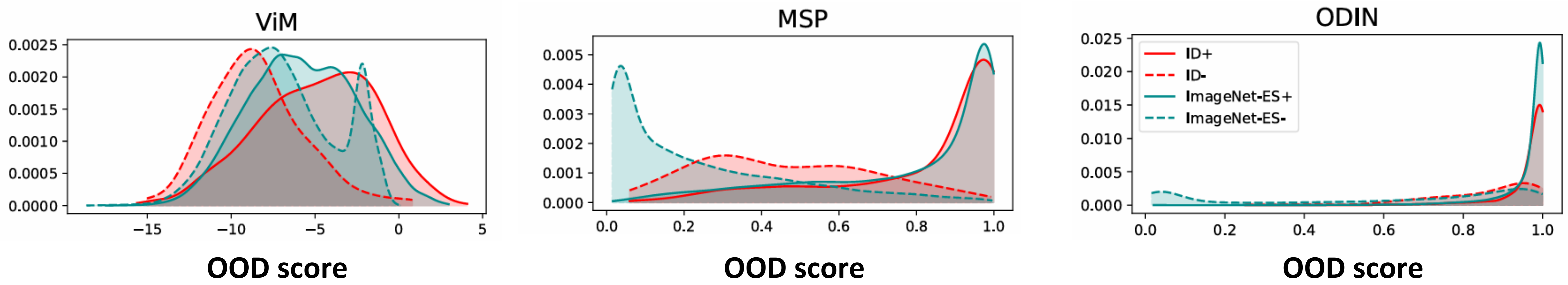


Experiments: Out-of-Distribution (OOD) detection

- Alternative framework: Model-Specific OOD framework
  - Model specific acceptance or rejection (MS-A or MS-R)
  - MS-A: correctly classified by model (ID+, C-OOD-)
  - MS-R: misclassified by model (S-OOD, ID-, C-OOD-)

Experiments: Out-of-Distribution (OOD) detection

- Alternative framework: Model-Specific OOD framework
  - Test on three OOD detection techniques: ViM, MSP, ODIN
  - Model: EfficientNet-B0
  - Two datasets
    - In-Distribution (ID): Tiny-ImageNet
    - Covariate shifted OOD (C-OOD): *ImageNet-ES*

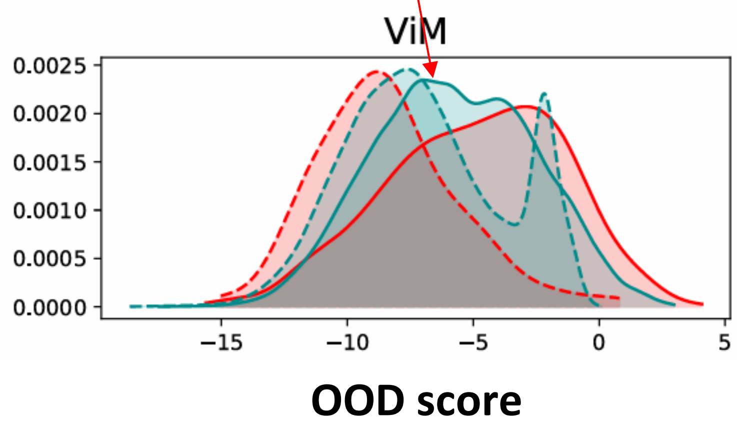


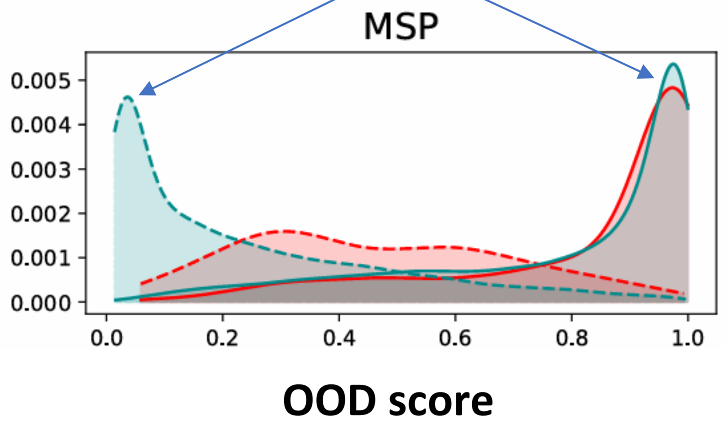
Experiments: Out-of-Distribution (OOD) detection

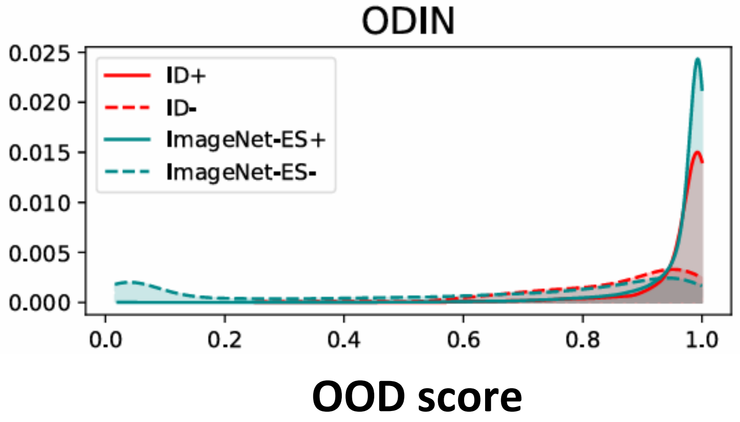
- Alternative framework: Model-Specific OOD framework
  - Test on three OOD detection techniques: ViM, MSP, ODIN
  - Model: EfficientNet-B0
  - Two datasets
    - In-Distribution (ID): Tiny-ImageNet
    - Covariate shifted OOD (C-OOD): *ImageNet-ES*

Still, no clear distinction between *ImageNet-ES+* and *ImageNet-ES-*

Clear separation between ImageNet-ES+ and ImageNet-ES-







Experiments: Out-of-Distribution (OOD) detection

- Evaluation of OOD detection methods
  - Do current OOD methods work consistently on real covariate shift samples?

[Consistency] OOD score - Classification Accuracy
(EfficientNet-B0)

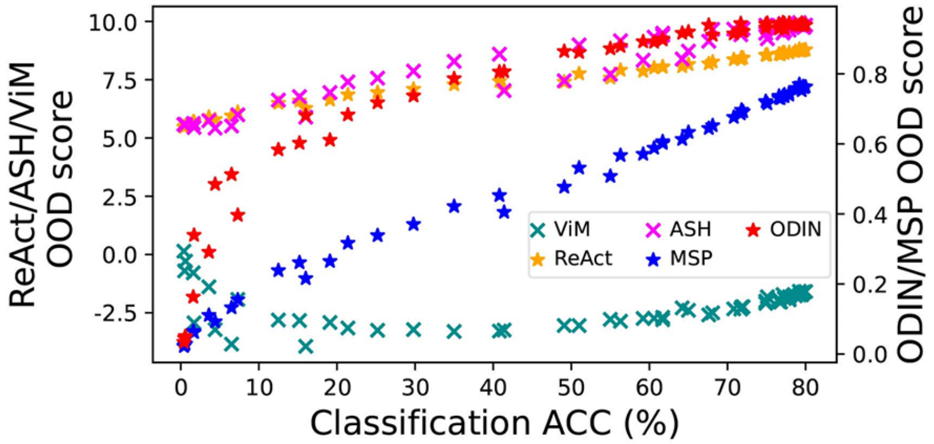


Figure 6. Each point represents the OOD score measured on the single parameter setting of *ImageNet-ES*.

Classical methods (MSP or ODIN) show more desirable correlation

SOTA method (ViM) accepts numerous samples as ID which are misclassified by the model

Experiments: Out-of-Distribution (OOD) detection

- Evaluation of OOD detection methods
  - Do current OOD methods work consistently on real covariate shift samples?

[Consistency] OOD score - Classification Accuracy
(EfficientNet-B0)

With ImageNet-ES, we found that no single method is superior in both C-OOD and S-OOD detection.

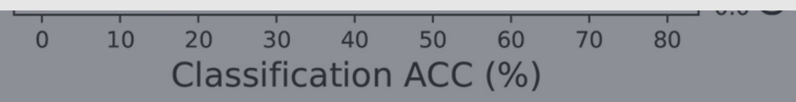


Figure 6. Each point represents the OOD score measured on the single parameter setting of *ImageNet-ES*.

samples as ID which are misclassified by the model

Experiments: Domain generalization

- How to enhance the robustness in the environmental and sensor domain (*ImageNet-ES*)?
  - **Basic** digital augmentation: color-jitter, solarize and posterize
  - **Advanced** digital augmentation: DeepAugment and AugMix
  - Include **real-world perturbed data** (*ImageNet-ES*) for finetuning

Experiments: Domain generalization

- How to enhance the robustness in the environmental and sensor domain (*ImageNet-ES*)?
  - **Basic** digital augmentation: color-jitter, solarize and posterize
  - **Advanced** digital augmentation: DeepAugment and AugMix
  - Include **real-world perturbed data** (*ImageNet-ES*) for finetuning

Table 2. Evaluation with different robustness enhancing strategies. The result is based on ResNet-50. (IN: ImageNet)

ID	Comp.aug	Basic digital aug	Advanced digital aug	Incl. *ImageNet-ES*	Eval dataset IN	IN-C	*ImageNet-ES*
1	✓				85.8	51.0	49.6
2	✓	✓			85.8	51.7	50.4
3	✓	✓	✓		85.5	57.4	49.1
4							
5							
6							

Digital augmentation improves the robustness on digitally corrupted images(ImageNet-C),

But NOT on real-world perturbed images.

Experiments: Domain generalization

- How to enhance the robustness in the environmental and sensor domain (*ImageNet-ES*)?
  - **Basic** digital augmentation: color-jitter, solarize and posterize
  - **Advanced** digital augmentation: DeepAugment and AugMix
  - Include **real-world perturbed data** (*ImageNet-ES*) for finetuning

Table 2. Evaluation with different robustness enhancing strategies. The result is based on ResNet-50. (IN: ImageNet)

ID	Comp.aug	Basic digital aug	Advanced digital aug	Incl. *ImageNet-ES*	Eval dataset IN	IN-C	*ImageNet-ES*
1	✓				85.8	51.0	49.6
2	✓	✓			85.8	51.7	50.4
3	✓	✓	✓		85.5	57.4	49.1
4	✓			✓	**86.0**	51.8	**55.8**
5	✓	✓		✓	85.8	51.4	54.5
6	✓	✓	✓	✓	84.0	**57.9**	53.7

Including ImageNet-ES data for finetuning improves the robustness on both digitally or real-world corrupted images.

Experiments: Sensor parameter control

- In practice, sensor parameter control is as important as obtaining smart model.

Table 3. Evaluation of various models on *ImageNet-ES*. (IN: ImageNet, AE: Auto exposure)

Model	Num. Params	Pretraining Dataset	DG method	IN	*ImageNet-ES* AE	All params	Best
ResNet-50 [8]	26M	IN-1K	-	86.3	32.2	50.2	80.1
		IN-21K	DeepAugment [13] +AugMix [12]	87.0	53.3	61.4	84.0
ResNet-152 [8]	60M	IN-1K	-	87.6	41.1	54.3	83.3
Efficientnet-B0 [32]	5M	IN-1K	-	88.1	51.4	58.1	83.8
Efficientnet-B3 [32]	12M	IN-1K	-	88.3	62.0	66.2	86.8
SwinV2-T [23]	28M	IN-1K	-	90.7	54.2	63.1	86.8
SwinV2-B [23]	88M	IN-1K	-	92.0	60.1	65.6	89.0
OpenCLIP-b [17]	87M	LAION-2B	Text-guided pretrain	94.3	66.3	71.0	92.7
OpenCLIP-h [17]	632M	LAION-2B	Text-guided pretrain	94.7	79.1	77.6	94.7
DINOv2-b [26]	90M	LVD-142M	Dataset curation	93.6	74.5	73.9	92.2
DINOv2-g [26]	1.1B	LVD-142M	Dataset curation	94.7	84.3	79.6	94.2

Well-tuned parameter setting (Best) improves the prediction accuracy
by 9.9 ~ 47.9 (vs Auto Exposure)
by 14.6 ~ 29.9 (vs All params)

Experiments: Sensor parameter control

- In practice, sensor parameter control is as important as obtaining smart model.

Table 3. Evaluation of various models on *ImageNet-ES*. (IN: ImageNet, AE: Auto exposure)

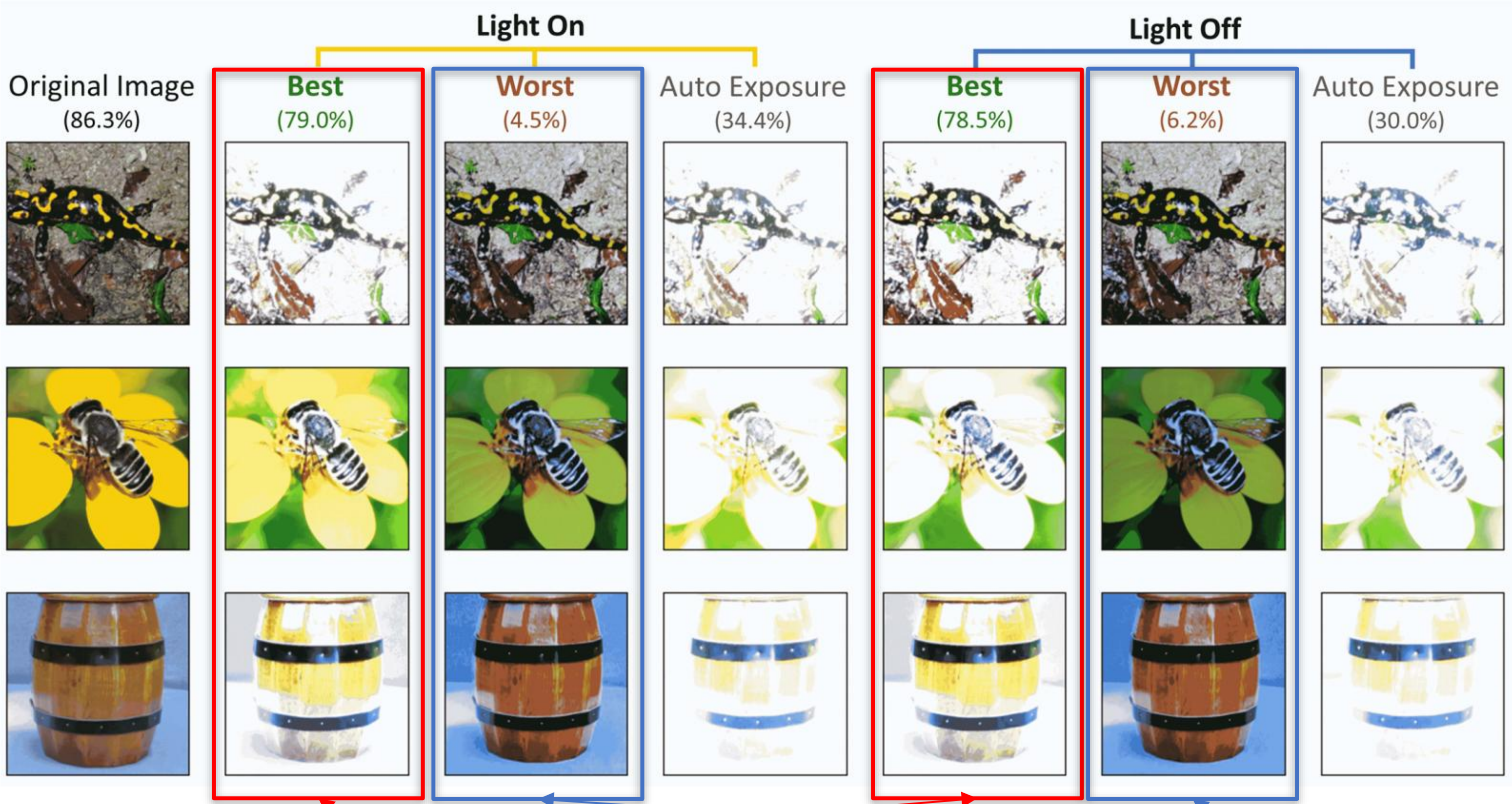
Model	Num. Params	Pretraining Dataset	DG method	IN	*ImageNet-ES* AE	*ImageNet-ES* All params	*ImageNet-ES* Best
ResNet-50 [8]	26M	IN-1K	-	86.3	32.2	50.2	80.1
		IN-21K	DeepAugment [13] +AugMix [12]	87.0	53.3	61.4	84.0
ResNet-152 [8]	60M	IN-1K	-	87.6	41.1	54.3	83.3
Efficientnet-B0 [32]	5M	IN-1K	-	88.1	51.4	58.1	83.8
Efficientnet-B3 [32]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	66.2	86.8
SwinV2-T [23]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	86.8
SwinV2-B [23]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	89.0
OpenCLIP-b [17]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	71.0	92.7
OpenCLIP-h [17]	632M	LAION-2B	Text-guided pretrain	94.7	79.1	77.6	94.7
DINOv2-b [26]	90M	LVD-142M	Dataset curation	93.6	74.5	73.9	92.2
DINOv2-g [26]	1.1B	LVD-142M	Dataset curation	94.7	84.3	79.6	94.2

120x larger model size
400x more training data

Even Efficientnet-B0 with the best outperforms OpenCLIP-h with auto exposure setting!

Experiments: Sensor parameter control

- Qualitative analysis on *ImageNet-ES*

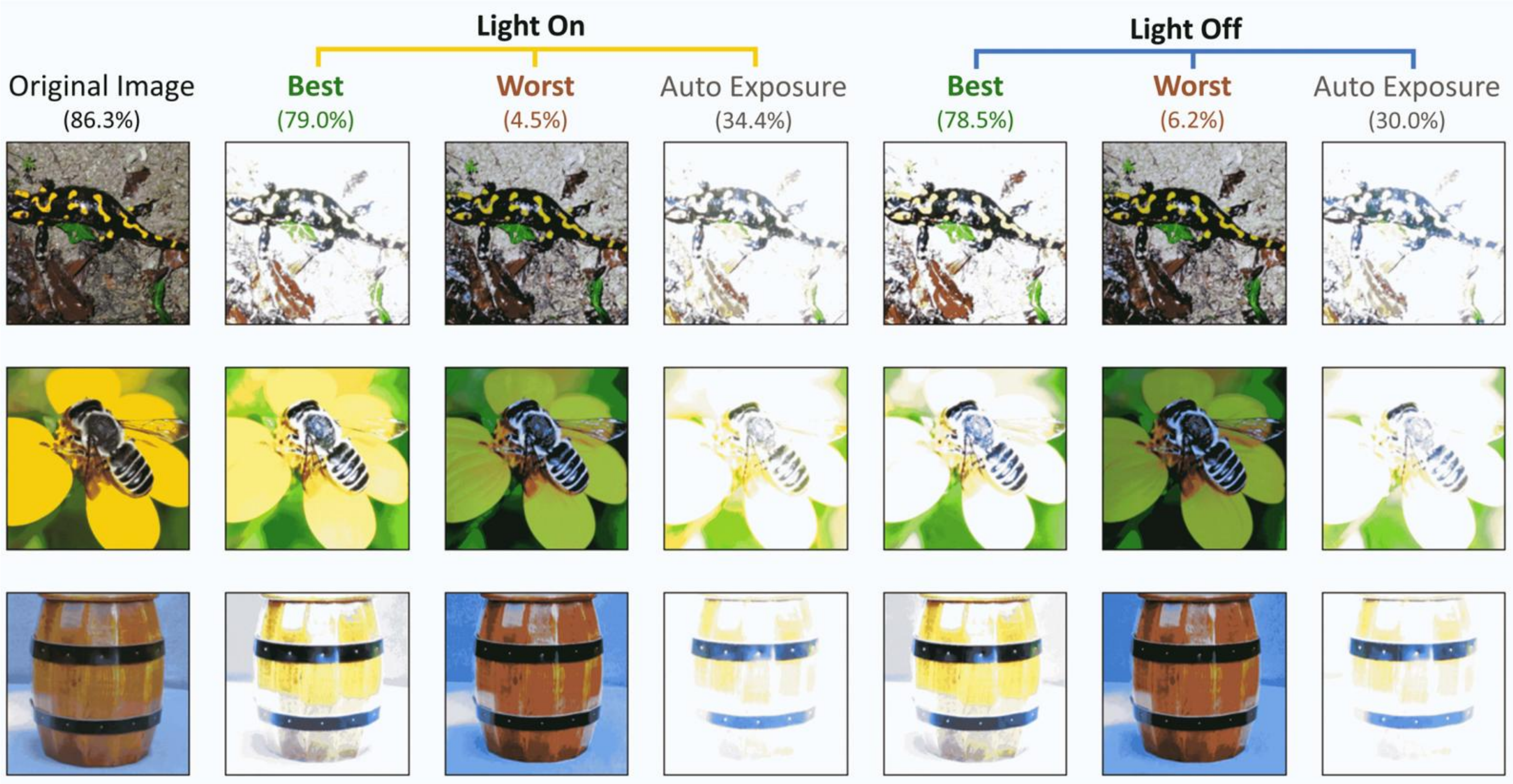


Looks too bright on human eyes
But model prefers them

Looks okay to human eyes,
Not to the model

Experiments: Sensor parameter control

- Qualitative analysis on *ImageNet-ES*



Sensor control should prioritize features based on **model's perspective**, rather than human intuition

Conclusion & Future work

- Investigated distribution shifts resulting from perturbations in Environmental and Sensor domains.
- ***ES-Studio**:* controllable testbed for environmental and sensor domains
- ***ImageNet-ES***: A novel covariate shifted dataset from the environment & sensor domain
- **OOD detection**: Limitation of semantics-centric framework => Need for new OOD detection method to incorporate both S-OOD and C-OOD
- **Domain generalization**: ES-augmentation improves the robustness in both conventional and ImageNet-ES benchmarks.
- **Sensor parameter control**
  - With well-tuned sensor parameters, light model could perform comparably to heavier and advanced model.
  - Need of model-centric design instead of relying solely on human aesthetics.
- Future work: Improve ES-Studio to take photos of real objects or printed photos, rather than capturing display.

Thank you:)