DisCoPatch: Taming Adversarially-driven Batch Statistics for Improved Out-of-Distribution Detection

Supplementary Material

The supplementary material is organized as follows: Appendix A covers the datasets used in this work. Appendix B describes the implementation details of DisCoPatch while Appendix C covers the compute resources required for training and evaluating the models. Appendix D presents the results of the experiments on Batch normalization bias in conventional setups. Appendix E covers the impact of alternative normalization layers in the discriminator performance. Appendix F analyses in more detail the effect that multiple patch counts have on DisCoPatch's performance. Appendix G provides detailed results on the ImageNet-1K Covariate Shift OOD detection benchmark.

A. Data Availability

ImageNet-1K [51] contains 1000 classes. For the Near-OOD experiments in this paper, we have employed the SSB-hard [59] and NINCO [3] datasets. In the case of Far-OOD, we have used iNaturalist [22], DTD [6], and OpenImage-O [61]. The experiments in the Covariate Shift section are evaluated in the ImageNet-1K(-C) [18] dataset. The OOD benchmark used to evaluate and compare the selected models closely follows the one proposed in OpenOOD by [70]. The images are resized to 256×256 before being fed to DisCoPatch.

The dataset ImageNet-1K(-C) was downloaded from its source [2]. Additionally, we have also used the original and publicly available splits for ImageNet-1K [3]. The remaining datasets and files containing training and evaluation splits were downloaded from OpenOOD's publicly available repository [4]. For convenience, DisCoPatch's repository includes the split files and a script that automatically downloads these datasets.

B. Implementation Details

DisCoPatch (our proposed method) is an Adversarial VAE, which is composed of a VAE and a Discriminator. The VAE features an Encoder ($\mathcal{E}_{\theta_\text{E}}$), consisting of convolutional layers with a kernel size of 3, stride 2, padding 1, and output padding of 1. All the convolution layers are followed by BN and a `LeakyReLU` activation function. The number of filters doubles with each layer. Encoded features are then flattened and passed through two distinct fully connected layers, one estimating z_μ and the other z_σ, with outputs the size of the latent dimension. These outputs undergo the reparametrization trick to generate z, which is then fed into the VAE's decoder, referred to as the Generator ($\mathcal{G}_{\theta_\text{G}}$). The Generator comprises transposed convolutions, followed by BN and a `LeakyReLU` activation, with the same kernel size, stride, padding, and output padding as the Encoder. The number of filters halves after each layer. A final convolutional layer with a kernel size of 3 and padding of 1, followed by a `Tanh` activation, generates the final output image. The generated image is subsequently fed into a Discriminator ($\mathcal{D}_\phi$). The Discriminator shares the same architecture as the Encoder but replaces the two fully connected layers with a single one that generates an output of size 1, followed by a `Sigmoid` activation. Additionally, for its recommended setup, `track_running_stats` is set to `False` in the Discriminator. The training process is covered in detail in Subsection 3.2, but can be summarized by Algorithm 1.

Algorithm 1 Training algorithm of DisCoPatch.

Initialize parameters of models θ, ϕ
while training **do**
 $x^{real} \leftarrow$ patches of images from dataset
 $z_\mu^{real}, z_\sigma^{real} \leftarrow \mathcal{E}_{\theta_\text{E}}(x^{real})$
 $z^{real} \leftarrow z_\mu^{real} + \epsilon_{real} z_\sigma^{real}$ with $\epsilon_{real} \sim \mathcal{N}(0, \mathbf{I})$
 $x^{rec} \leftarrow \mathcal{G}_{\theta_\text{G}}(z^{real})$
 $z^{fake} \leftarrow \epsilon_{fake}$ with $\epsilon_{fake} \sim \mathcal{N}(0, \mathbf{I})$
 $x^{fake} \leftarrow \mathcal{G}_{\theta_\text{G}}(z^{fake})$
 $x^{rec} \leftarrow \mathcal{G}_{\theta_\text{G}}(z^{real})$
 $D^{real} \leftarrow \mathcal{D}_\phi(x^{real})$
 $D^{rec}, D^{fake} \leftarrow \mathcal{D}_\phi(x^{rec}), \mathcal{D}_\phi(x^{fake})$

 $\theta \xleftarrow{} \nabla_\theta \mathcal{L}_\text{VAE}(\theta)$
 $\phi \xleftarrow{} \nabla_\phi \mathcal{L}_\text{D}(\phi)$
end while

Table 2. Hyperparameters used for DisCoPatch's training.

Model	Lat. Dim.	Hidden Dimensions	lr
Full-Size	1024	32, 64, 128, 256, 512, 1024	$5e^{-5}$
Patches	1024	128, 256, 512, 1024	$8.5e^{-5}$

DisCoPatch is optimized using the Adam optimizer, with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. Both models share the same learning rate, lr. As shown in Equation 6, three weighing

[2] https://github.com/hendrycks/robustness
[3] https://huggingface.co/datasets/benjamin-paine/imagenet-1k-256x256
[4] https://github.com/jingkang50/openood

terms are required to train the model; these were fixed for all datasets, with $\omega_{\text{KL}} = 1e^{-4}$, $\omega_{\text{Rec}} = 1e^{-3}$ and $\omega_{\text{Gen}} = 1e^{-3}$. Additional hyperparameters can be found in Table 2. The developed code is based on a publicly available repository [5]. The referred repository is released under the Apache 2.0 License.

C. Compute Resources

This appendix describes the computational resources employed for inference in the selected models and to train DisCoPatch.

C.1. Training

DisCoPatch was trained on ImageNet-1K using a system equipped with an NVIDIA H100 Tensor Core GPU (94 GB VRAM), a 32-core, 64-thread AMD EPYC 9334 CPU, and 768 GB RAM. More information can be found in Table 3.

Table 3. Summary of the compute resources required for training the DisCoPatch models on ImageNet-1K.

Model	#Patches	Batch Size	Epochs	Parameters	Time (s)
Full-Size	1	660	70	69,240,517	111,471
Patches	48	67	30	69,118,340	94,323

C.2. Inference

To measure the models' latency, we fed them 1000 individual inputs and calculated the average inference time per image. Table 4 reveals DisCoPatch has the lowest latency. Although MobileNetV2 is smaller, it takes 224×224 inputs, whereas DisCoPatch processes an image as a batch of 64×64 patches. Latency was measured on a machine with an NVIDIA RTX4070 GPU (8 GB VRAM), an 8-core, 16-thread AMD RYZEN 9 8945HS CPU, and 32 GB RAM.

Table 4. Latency of the tested models.

Model	#Parameters	Latency (ms)
MOODv2 (BEiTv2)	86,530,984	19.26
SCALE (ResNet-50)	25,557,032	11.27
NNGuide (RegNet)	83,590,140	31.00
NNGuide (ResNet-50)	25,557,032	11.10
NNGuide (MobileNetv2)	**3,504,872**	9.40
RankFeat (ResNetv2-101)	44,549,160	15.05
ASH (ResNet-50)	25,557,032	11.15
FDBD (ResNet-50)	25,557,032	11.17
DisCoPatch-64	6,218,753	**1.56**

D. Batch Normalization Bias

Table 5 reveals a critical limitation in the model's behavior when evaluated with the BatchNorm employed in its standard *evaluation mode*: the model fails to distinguish between ID and OOD samples reliably. In contrast, when disabling the use of the learned statistics and instead using batch-specific statistics, the model's performance improves significantly, even with a batch size of 1. This effect demonstrates that the running statistics acquired during training are ineffective for discriminating ID from OOD, while the *test* batch statistics provide more discriminating power for detecting OOD samples. It should be noted that as the batch size increases, this improvement becomes more pronounced, which indicates that the model has developed a dependency/shortcut on batch-specific statistics, instead of leveraging the running mean and variance acquired during training. This means that the use of BatchNorm's running statistics compromises robustness, as it has been observed in adversarial and OOD scenarios [1, 62].

Table 5. OOD detection performance, reported as AUROC/FPR95, of a DisCoPatch model trained on complete ImageNet-1k images. Legend: BS = Batch Size.

Mode		Near-OOD	Far-OOD
Learned Statistics		38.4/98.0	34.5/98.1
Batch Statistics	BS=1	64.9/87.4	70.9/71.2
	BS=16	90.2/55.5	91.1/40.0
	BS=32	95.7/28.6	93.4/36.2
	BS=64	99.3/2.2	96.1/23.8
	BS=128	99.8/0.3	97.4/17.0
	BS=256	**100.0/0.0**	**98.1/12.5**

It is important to note that in the configuration employed for this experiment, each batch contains exclusively ID or OOD samples. This means that a single anomaly score predicted for an image by the *Batch Statistics* mode is dependent on the statistics from every image in the batch. This design constraint limits the suitability of this configuration for multiple applications because it requires that all images in a batch share the same class type. A practical and effective solution to ensure this homogeneity without prior class knowledge is by constructing each batch from patches of the same image.

E. Alternative Normalization Layers

To better demonstrate the effect of the BatchNorm layer and batch statistics, we replaced BatchNorm with GroupNorm and InstanceNorm and retrained DisCoPatch. As shown in Table 6, GroupNorm performs the worst, while InstanceNorm shows intermediate results. These findings suggest that avoiding normalization across multiple channels is beneficial. However, BatchNorm consistently achieves the best performance, particularly under Covariate Shift, highlighting the importance of batch-level statistics.

[5] https://github.com/AntixK/PyTorch-VAE

Table 6. Comparison between different normalization layers on DisCoPatch. The results correspond to the 64-patch configuration.

Normalization	Near-OOD	Far-OOD	Covariate Shift OOD
GroupNorm	85.0	81.0	78.5
InstanceNorm	92.8	93.7	86.2
BatchNorm	**95.1**$^{+2.3}$	**96.6**$^{+2.9}$	**97.2**$^{+11.0}$

F. Patch Count Imapct

Table 7 shows that the use of more patches improves the detection accuracy, reaching a plateau around 64 patches.

Table 7. OOD detection results on ImageNet-1K for multiple patch counts.

Model	Near-OOD	Far-OOD	Covar. Shift OOD
MOODv2 (BEiTv2)	88.9	97.1	70.5
SCALE (ResNet-50)	81.4	96.5	83.3
NNGuide (RegNet)	89.2	**97.8**	78.5
RankFeat [57] (ResNetv2-101)	89.7	94.5	91.9
ASH [7] (ResNet-50)	78.6	96.1	84.7
FDBD [29] (ResNet-50)	74.9	95.8	82.2
DisCoPatch-4 (Proposed)	90.3$^{+0.6}$	92.6$^{-5.2}$	93.0$^{+1.1}$
DisCoPatch-16 (Proposed)	94.3$^{+4.6}$	96.0$^{-1.8}$	96.5$^{+3.6}$
DisCoPatch-32 (Proposed)	94.8$^{+5.1}$	96.4$^{-1.4}$	96.9$^{+4.0}$
DisCoPatch-64 (Proposed)	95.1$^{+5.4}$	96.6$^{-1.2}$	97.2$^{+4.3}$
DisCoPatch-128 (Proposed)	95.2$^{+5.5}$	96.7$^{-1.1}$	97.3$^{+4.4}$
DisCoPatch-512 (Proposed)	**95.3**$^{+5.6}$	96.8$^{-1.0}$	**97.4**$^{+4.5}$

G. Covariate Shift Results on ImageNet-1K

This appendix contains the performance metrics per corruption achieved on the Covariate Shift OOD benchmark.

G.1. MOODv2 (BEiTv2)

MOODv2 obtains its best results for corruptions that filter high-frequency components. Table 8 shows that it scores very low for Intensity 1 in all corruption tests.

Table 8. Covariate shift OOD benchmark for MOODv2.

Corruption	Corruption Intensity					Average
	1	2	3	4	5	
Brightness	54.2/93.8	55.2/93.2	56.9/92.2	59.2/90.6	62.2/88.2	57.5/91.6
Contrast	59.6/91.6	61.7/90.4	66.1/87.7	77.5/76.6	86.4/54.9	70.3/80.2
Defocus Blur	69.6/80.3	76.0/70.7	85.5/51.0	91.8/33.9	95.4/21.0	83.7/51.4
Elastic Transform	60.2/88.5	75.9/62.7	63.2/85.1	70.8/75.3	87.1/44.2	71.4/71.1
Fog	70.0/82.5	77.3/71.5	89.2/40.0	93.3/25.6	97.0/11.4	85.3/46.2
Frost	61.8/88.8	70.3/79.9	75.6/71.7	77.1/69.7	80.0/63.5	72.9/74.7
Glass Blur	60.6/90.6	72.1/78.0	81.6/62.0	88.9/44.4	96.5/17.6	79.9/58.5
Gaussian Blur	58.0/90.0	60.0/87.7	64.8/82.8	71.5/74.1	80.3/59.0	66.9/78.7
Gaussian Noise	63.2/86.3	70.2/77.8	83.9/51.0	87.9/41.4	93.3/26.5	79.7/56.6
Impulse Noise	57.2/89.8	60.4/86.8	63.6/83.5	71.0/74.6	78.9/61.6	66.2/79.3
JPEG Compression	63.5/88.0	65.9/85.7	67.6/83.8	71.9/77.6	77.5/68.5	69.2/80.7
Motion Blur	58.7/90.2	63.1/85.9	70.5/77.0	80.0/61.0	85.8/48.5	71.6/72.5
Pixelate	55.8/92.2	57.5/90.9	61.0/87.9	67.7/81.0	83.8/57.1	65.2/81.8
Saturate	54.0/93.4	55.6/92.2	55.4/93.1	60.3/90.0	65.2/85.2	58.1/90.8
Shot Noise	58.3/89.7	61.0/86.8	65.3/82.1	73.7/70.7	80.1/59.1	67.7/77.7
Snow	62.3/87.4	70.8/77.8	70.7/78.9	75.6/71.4	77.2/67.7	71.3/76.6
Spatter	55.3/92.9	59.1/90.2	62.0/87.8	64.2/85.7	69.4/80.0	62.0/87.3
Speckle Noise	57.6/90.4	59.1/88.8	63.9/83.7	67.4/79.3	72.3/72.0	64.1/82.8
Zoom Blur	65.8/83.7	71.9/75.5	76.7/67.1	81.3/58.3	86.5/46.2	76.5/66.1
Average	60.3/88.9	65.4/82.8	69.7/76.2	75.3/67.4	81.8/54.3	70.5/73.9

G.2. SCALE (ResNet-50)

The results in Table 9 demonstrate a drop in performance for Intensity 2 in some corruptions that dampen the high-frequency components, such as blurs. This occurs despite good performance on intensity 1.

Table 9. Covariate shift OOD benchmark for SCALE.

Corruption	Corruption Intensity					Average
	1	2	3	4	5	
Brightness	60.1/90.3	61.2/89.7	63.9/88.0	68.3/84.8	74.1/79.0	65.5/86.4
Contrast	74.5/76.9	80.3/67.7	88.6/48.6	98.1/9.6	99.9/0.3	88.3/40.6
Defocus Blur	82.3/63.6	87.3/52.4	93.9/31.5	97.1/15.5	98.8/6.1	91.9/33.8
Elastic Transform	71.1/80.3	82.5/62.5	81.3/62.9	87.9/46.7	95.0/19.8	83.6/54.4
Fog	83.4/53.9	73.1/79.0	77.6/72.7	83.2/62.3	86.9/53.1	80.8/64.2
Frost	93.7/29.9	72.5/80.0	85.2/59.6	91.3/41.1	92.2/38.0	87.0/49.7
Gaussian Blur	94.7/26.8	74.6/76.3	85.0/57.8	92.3/36.9	96.1/20.9	88.6/43.7
Gaussian Noise	99.1/4.3	66.8/86.4	76.0/77.3	87.7/54.1	95.8/21.9	85.1/48.8
Glass Blur	99.3/3.5	82.4/62.0	89.9/43.1	97.2/14.7	98.2/9.4	93.4/26.5
Impulse Noise	99.0/4.7	75.6/77.7	83.2/65.6	89.0/50.0	96.7/16.9	88.7/43.0
JPEG Compression	99.2/3.5	70.7/81.0	73.0/78.6	74.8/76.6	81.3/67.8	79.8/61.5
Motion Blur	89.1/48.9	73.7/78.0	81.6/66.5	90.9/43.5	96.6/18.2	86.4/51.0
Pixelate	98.3/8.3	65.0/86.9	64.9/87.1	79.5/70.6	89.6/46.3	79.4/59.8
Saturate	93.4/32.2	65.3/86.9	64.1/88.6	61.4/89.6	68.4/85.2	70.5/76.5
Shot Noise	76.0/77.3	68.0/85.2	77.8/74.1	88.0/52.3	96.5/17.8	81.3/61.3
Snow	98.7/6.5	73.0/80.2	88.5/50.2	85.8/59.7	91.5/40.8	87.5/47.5
Spatter	95.2/24.7	61.5/89.4	69.2/83.6	76.3/75.5	79.0/70.4	76.2/68.7
Speckle Noise	84.1/58.1	66.8/85.9	70.9/82.3	83.7/63.0	89.5/47.3	79.0/67.3
Zoom Blur	94.2/29.0	81.9/64.0	87.3/51.2	90.4/42.3	93.0/32.9	89.3/43.9
Average	88.7/38.0	72.7/77.4	79.0/66.8	85.4/52.0	90.5/36.4	83.3/54.1

G.3. NNGuide (RegNet)

NNGuide surpasses the performance of MOODv2, as demonstrated by the results in Table 10, particularly for higher corruption intensities. Nonetheless, it also suffers from significantly low scores at Intensity 1.

Table 10. Covariate shift OOD benchmark for NNGuide (RegNet).

Corruption	Corruption Intensity					Average
	1	2	3	4	5	
Brightness	57.9/91.7	59.8/90.9	62.8/89.2	67.3/85.7	73.0/79.8	64.2/87.5
Contrast	75.8/76.1	82.3/64.5	91.4/37.9	98.7/6.1	99.9/0.4	89.6/37.0
Defocus Blur	72.1/75.6	78.1/66.0	87.9/43.9	93.7/26.2	96.9/14.2	85.7/45.2
Elastic Transform	66.2/84.8	78.0/66.5	73.4/75.0	82.7/57.8	94.6/23.5	79.0/61.5
Fog	75.9/75.9	81.2/66.3	88.2/48.5	91.7/37.3	96.1/19.6	86.6/49.5
Frost	70.8/81.4	81.7/64.0	87.8/49.1	88.7/46.7	91.6/37.0	84.1/55.6
Gaussian Blur	64.0/86.4	75.0/71.3	83.9/54.6	90.0/38.8	96.7/15.2	81.9/53.3
Gaussian Noise	64.0/86.6	70.3/79.9	80.5/63.8	90.8/37.6	98.0/9.5	80.7/55.5
Glass Blur	71.2/77.9	80.3/62.8	93.4/27.2	95.7/18.6	97.4/11.5	87.6/39.6
Impulse Noise	62.2/87.3	68.2/80.9	74.2/72.8	87.4/46.2	96.6/15.7	77.7/60.6
JPEG Compression	60.9/89.0	63.7/86.7	65.9/84.9	71.7/78.3	78.8/68.0	68.2/81.4
Motion Blur	62.6/86.7	69.0/79.6	79.3/63.5	89.4/39.7	93.9/25.3	78.8/59.0
Pixelate	62.3/88.5	63.8/87.1	68.1/82.4	75.4/72.1	82.1/60.2	70.3/78.0
Saturate	61.3/89.7	62.9/88.2	58.5/91.6	68.7/84.7	77.7/72.1	65.8/85.2
Shot Noise	66.0/85.1	72.9/76.8	81.1/62.0	92.7/30.6	97.3/12.3	82.0/53.4
Snow	69.6/82.5	80.9/64.4	80.1/65.5	86.4/51.1	91.0/37.5	81.6/60.2
Spatter	62.1/89.4	67.7/85.7	70.5/82.3	74.7/78.2	81.6/66.4	71.3/80.4
Speckle Noise	63.8/87.4	66.8/84.4	76.6/70.4	82.5/58.4	89.2/41.3	75.8/68.4
Zoom Blur	70.1/79.0	76.7/69.1	82.7/57.8	86.5/48.6	90.4/38.0	81.3/58.5
Average	66.3/84.3	72.6/75.5	78.2/64.3	85.0/49.6	90.7/34.1	78.5/61.6

G.4. NNGuide (ResNet-50)

The behavior observed for this backbone of NNGuide in Table 11 is very similar to the one observed when the RegNet was used. However, it is slightly more effective than the bigger backbone at Covariate Shift detection.

Table 11. Covariate shift OOD benchmark for NNGuide (ResNet-50).

Corruption	Corruption Intensity					Average
	1	2	3	4	5	
Brightness	56.5/90.0	58.8/88.5	62.7/85.6	68.3/80.5	74.7/72.6	64.2/83.4
Contrast	54.1/91.3	61.9/86.2	75.0/72.1	93.3/27.8	99.2/3.0	76.7/56.1
Defocus Blur	74.7/69.2	81.2/57.6	90.3/36.5	94.7/22.6	97.2/13.0	87.6/39.8
Elastic Transform	64.8/83.3	80.4/65.4	77.6/66.9	86.4/51.0	95.7/22.1	81.0/57.7
Fog	69.5/79.0	74.6/73.2	80.7/63.9	84.5/55.3	92.2/32.8	80.3/60.8
Frost	71.7/73.9	84.8/49.8	91.0/33.4	91.7/31.0	94.2/22.7	86.7/42.2
Gaussian Blur	50.3/93.5	65.9/81.8	77.6/67.9	85.9/51.7	94.7/23.6	74.9/63.7
Gaussian Noise	66.2/82.5	74.4/73.1	84.9/53.6	93.8/27.1	98.6/6.4	83.6/48.5
Glass Blur	76.6/68.1	86.4/47.7	96.1/16.5	97.6/10.5	98.6/6.2	91.1/29.8
Impulse Noise	78.1/67.9	82.3/61.0	85.9/52.5	93.8/27.6	98.3/8.2	87.7/43.5
JPEG Compression	63.4/84.7	66.6/81.5	69.3/78.4	77.6/66.5	87.2/46.5	72.8/71.5
Motion Blur	69.3/76.6	78.8/62.3	88.7/41.2	94.9/22.1	97.0/14.0	85.7/43.2
Pixelate	63.3/86.5	65.2/84.9	75.6/71.9	86.9/48.2	92.2/32.1	76.7/64.7
Saturate	47.3/96.3	49.9/95.2	45.5/96.9	57.2/94.6	66.5/91.2	53.3/94.8
Shot Noise	68.1/80.7	77.2/69.5	86.1/51.7	95.0/23.1	97.9/10.3	84.9/47.1
Snow	74.9/74.5	89.2/42.3	87.6/48.4	92.8/31.5	95.2/21.5	87.9/43.7
Spatter	43.5/97.1	55.6/95.7	65.0/93.6	71.2/91.4	78.2/87.1	62.7/93.0
Speckle Noise	52.0/95.1	57.2/93.2	71.5/84.2	78.6/75.7	85.4/62.5	68.9/82.2
Zoom Blur	77.2/68.8	83.4/57.8	87.6/47.8	90.5/39.5	92.9/31.3	86.3/49.0
Average	64.3/82.0	72.3/71.9	78.9/61.2	86.0/46.2	91.4/32.0	78.6/58.7

G.5. NNGuide (MobileNet)

The behavior observed for this backbone of NNGuide in Table 12 is very similar to the one observed when the RegNet and ResNet were used. Its performance is lower than the one achieved by the other backbones but by a small margin.

Table 12. Covariate shift OOD benchmark for NNGuide (MobileNet).

Corruption	Corruption Intensity					Average
	1	2	3	4	5	
Brightness	52.6/94.1	54.0/93.6	56.6/92.7	60.8/90.7	66.0/87.2	58.0/91.7
Contrast	55.9/93.0	58.3/91.7	62.9/88.2	76.5/70.7	91.6/32.5	69.1/75.2
Defocus Blur	68.2/82.8	76.3/71.5	88.6/44.7	94.4/24.5	96.8/14.7	84.9/47.6
Elastic Transform	58.2/92.1	73.6/79.3	70.5/84.0	80.8/71.2	90.9/48.9	74.8/75.1
Fog	59.1/92.1	62.6/90.2	69.5/84.0	75.9/74.6	87.1/51.2	70.8/78.4
Frost	63.8/89.7	76.5/77.6	83.2/66.6	84.4/64.0	87.4/56.6	79.0/70.9
Gaussian Blur	58.0/91.4	72.9/76.1	86.3/50.0	93.9/26.2	98.1/9.2	81.8/50.6
Gaussian Noise	68.0/82.9	80.1/63.4	93.3/27.2	98.6/6.1	99.8/0.9	88.0/36.1
Glass Blur	71.8/80.4	83.1/59.6	94.9/22.6	97.0/14.0	98.4/7.3	89.0/36.8
Impulse Noise	72.9/75.6	83.5/57.3	90.9/36.1	98.5/6.6	99.8/0.8	89.1/35.3
JPEG Compression	60.0/92.3	62.3/91.6	64.0/90.7	69.3/87.1	76.2/78.8	66.4/88.1
Motion Blur	62.4/88.6	73.6/75.8	86.6/48.3	94.7/21.9	97.0/13.0	82.9/49.5
Pixelate	61.5/91.1	67.2/86.7	76.1/73.3	90.2/37.2	94.2/22.3	77.9/62.1
Saturate	55.0/94.0	56.0/93.4	55.8/92.4	65.3/86.6	73.0/78.4	61.0/89.0
Shot Noise	68.7/82.5	82.1/59.6	93.1/28.3	98.7/6.0	99.6/1.8	88.4/35.6
Snow	71.7/80.7	85.4/59.1	82.6/65.2	88.1/52.1	90.5/45.6	83.7/60.5
Spatter	55.4/92.7	69.2/82.3	76.1/77.2	81.1/66.8	85.8/60.1	73.5/75.8
Speckle Noise	64.8/87.2	70.7/80.6	86.1/50.9	91.9/33.2	96.0/18.0	81.9/54.0
Zoom Blur	74.0/76.5	81.5/63.6	86.1/53.6	89.4/43.9	91.8/36.0	84.6/54.7
Average	63.3/87.3	72.0/76.5	79.1/61.9	85.8/46.5	90.5/34.9	78.1/61.4

G.6. RankFeat (ResNetv2-101)

The performance achieved by RankFeat in Table 13 is significantly superior to that achieved by the other baseline methods. Nevertheless, it remains inferior to DisCoPatch in terms of Covariate Shift OOD detection abilities.

Table 13. Covariate shift OOD benchmark for RankFeat.

Corruption	Corruption Intensity					Average
	1	2	3	4	5	
Brightness	80.9/67.3	82.6/64.2	84.7/60.0	87.1/54.5	89.6/47.9	85.0/58.8
Contrast	83.2/63.4	84.7/60.6	87.3/54.9	92.8/37.6	97.4/14.3	89.1/46.1
Defocus Blur	88.1/52.9	90.4/46.5	93.0/39.4	94.7/31.6	96.1/23.7	92.5/38.8
Elastic Transform	84.0/61.8	87.6/56.2	88.7/44.3	90.8/37.9	95.4/22.1	89.3/44.4
Fog	85.1/60.3	86.8/56.8	89.3/50.9	90.8/46.0	93.6/33.9	89.1/49.6
Frost	89.1/43.6	94.2/25.2	96.4/16.2	96.7/14.9	97.6/10.8	94.8/22.1
Gaussian Blur	83.7/62.6	89.1/51.0	91.9/43.8	93.7/38.0	96.3/22.3	91.0/43.5
Gaussian Noise	87.8/52.3	91.1/42.2	94.4/28.8	96.9/15.8	98.7/5.4	93.8/28.9
Glass Blur	90.9/38.7	93.9/27.9	97.1/14.5	97.4/13.4	97.0/16.9	95.3/22.3
Impulse Noise	94.0/31.3	95.4/24.9	96.2/20.2	97.7/11.2	98.8/4.7	96.4/18.5
JPEG Compression	84.7/61.7	86.6/57.9	87.9/55.0	91.1/45.8	93.5/37.1	88.7/51.5
Motion Blur	84.9/59.0	88.5/50.5	92.7/37.9	96.0/22.7	97.4/14.6	91.9/36.9
Pixelate	84.6/62.3	85.2/62.8	90.4/48.3	95.4/26.3	97.5/12.9	90.6/42.5
Saturate	81.4/67.6	81.2/68.8	82.6/64.9	88.4/52.1	91.1/43.4	84.9/59.3
Shot Noise	88.2/50.4	91.7/39.1	94.5/27.4	97.3/13.1	98.5/6.7	94.0/27.3
Snow	90.8/43.2	95.4/21.7	94.4/28.4	96.2/19.8	97.2/14.0	94.8/25.4
Spatter	81.5/67.1	86.1/60.1	90.7/46.0	93.1/36.9	95.3/25.3	89.4/47.2
Speckle Noise	87.2/52.4	89.2/46.7	93.4/31.8	95.0/24.9	96.4/18.0	92.2/34.8
Zoom Blur	88.7/51.1	91.1/44.5	93.0/37.0	94.2/32.1	95.5/25.6	92.5/38.1
Average	86.3/55.2	89.0/47.8	91.5/39.5	94.0/30.2	95.9/21.0	91.3/38.7

G.7. ASH (ResNet-50)

The behavior observed for ASH in Table 14 is very similar to the one observed in SCALE, which uses a similar backbone.

Table 14. Covariate shift OOD benchmark for ASH.

Corruption	Corruption Intensity					Average
	1	2	3	4	5	
Brightness	64.6/86.9	65.5/85.3	67.9/82.6	71.9/77.5	77.1/69.7	69.4/80.4
Contrast	79.2/68.6	84.7/57.4	92.2/35.9	99.0/4.9	100.0/0.1	91.0/33.4
Defocus Blur	83.4/60.8	88.1/48.2	94.3/27.0	97.2/14.1	98.7/6.4	92.4/31.3
Elastic Transform	72.0/79.4	79.1/69.9	81.8/65.5	87.6/52.7	95.2/27.6	83.1/59.0
Fog	75.5/72.8	86.9/50.3	92.5/34.0	93.2/41.6	95.3/23.1	88.7/44.4
Frost	75.5/72.8	86.9/50.3	92.5/34.0	93.2/31.6	95.3/23.1	88.7/42.4
Gaussian Blur	75.7/75.1	86.4/53.4	92.9/33.5	96.4/18.5	99.2/4.0	90.1/36.9
Gaussian Noise	70.4/79.3	77.8/69.9	88.1/48.5	95.7/21.7	99.1/3.9	86.2/44.7
Glass Blur	82.0/64.8	89.0/46.7	96.3/18.8	97.6/12.3	98.7/6.5	92.7/29.8
Impulse Noise	77.0/71.6	83.4/60.6	88.4/48.1	96.1/20.1	99.1/4.2	88.8/40.9
JPEG Compression	71.8/80.6	73.8/77.7	75.4/75.3	80.8/65.3	87.9/48.7	77.9/69.5
Motion Blur	75.8/72.7	82.7/59.5	90.8/38.4	96.2/18.6	98.0/10.0	88.7/39.8
Pixelate	68.7/84.1	70.2/82.3	78.7/70.5	88.5/48.5	93.1/32.6	79.9/63.6
Saturate	68.4/83.8	69.2/81.3	66.0/85.7	72.9/78.4	80.5/67.6	71.4/79.4
Shot Noise	71.1/78.9	79.8/66.9	88.7/47.3	96.5/18.3	98.6/7.0	86.9/43.7
Snow	75.1/75.3	89.4/45.0	86.1/52.7	92.1/34.9	95.5/21.9	87.6/45.9
Spatter	64.8/87.3	71.5/82.1	79.5/72.9	80.5/70.6	86.9/58.3	76.6/74.2
Speckle Noise	69.3/81.5	72.9/77.4	84.8/57.9	90.1/44.1	94.2/28.8	82.3/57.9
Zoom Blur	78.8/72.6	83.4/64.9	87.0/56.2	89.8/48.7	92.0/41.1	86.2/56.7
Average	73.6/76.2	80.0/64.7	85.5/51.8	90.3/38.0	93.9/25.5	84.7/51.3

G.8. FDBD (ResNet-50)

The behavior of FDBD presented in Table 15 is analogous to that seen in SCALE and ASH, with the three sharing the same backbone.

Table 15. Covariate shift OOD benchmark for FDBD.

Corruption	Corruption Intensity					Average
	1	2	3	4	5	
Brightness	62.1/89.2	62.7/88.5	64.4/86.6	67.8/83.3	72.7/77.6	65.9/85.0
Contrast	76.1/75.0	81.7/65.5	89.6/46.1	98.2/10.0	99.9/0.3	89.1/39.4
Defocus Blur	79.2/70.9	84.0/61.1	91.2/40.6	95.2/24.8	97.6/13.2	89.4/42.1
Elastic Transform	67.2/85.1	74.2/76.1	76.9/72.7	83.3/61.3	93.1/37.4	79.0/66.5
Fog	71.8/80.6	75.2/76.5	79.7/69.6	83.1/62.2	91.0/41.7	80.1/66.1
Frost	71.3/79.0	83.6/59.4	90.0/43.2	90.9/40.3	93.6/30.8	85.9/50.5
Gaussian Blur	72.2/81.0	82.4/65.0	90.2/45.2	94.7/27.6	98.5/7.5	87.6/45.3
Gaussian Noise	71.4/80.5	79.2/70.2	88.5/49.9	95.5/24.1	99.0/4.7	86.7/45.9
Glass Blur	77.8/72.5	85.0/57.3	94.0/28.0	95.7/20.9	97.3/13.6	90.0/38.4
Impulse Noise	76.4/74.2	83.5/63.2	88.4/50.9	95.8/22.8	98.9/5.1	88.6/43.3
JPEG Compression	70.8/83.0	72.9/80.3	74.4/78.2	79.2/70.6	85.6/56.9	76.6/73.8
Motion Blur	71.2/79.6	78.1/68.6	87.7/47.9	94.6/25.9	97.0/15.2	85.7/47.4
Pixelate	69.0/85.5	70.7/83.7	78.6/73.9	87.3/55.0	91.7/40.2	79.5/67.7
Saturate	65.1/86.7	66.5/84.6	62.3/89.2	69.1/83.4	78.1/72.5	68.2/83.3
Shot Noise	70.7/80.9	79.9/68.8	88.6/49.7	96.2/20.4	98.4/8.0	86.8/45.5
Snow	72.9/78.5	88.0/50.4	84.5/57.6	90.8/40.8	94.4/27.8	86.1/51.0
Spatter	61.9/89.7	68.6/84.6	77.7/75.2	78.7/72.6	85.9/60.2	74.6/76.5
Speckle Noise	66.7/84.3	70.9/80.3	83.7/61.5	89.2/47.9	93.6/32.3	80.8/61.3
Zoom Blur	72.9/81.1	77.5/76.0	81.5/69.6	84.8/63.7	87.6/56.7	80.9/69.4
Average	70.9/80.9	77.1/71.6	82.7/59.8	87.9/45.1	92.3/31.7	82.2/57.8

G.9. DisCoPatch

As seen in Table 16, DisCoPatch excels at detecting every sort of corruption at each possible intensity on ImageNet-1K.

Table 16. Covariate shift OOD benchmark for DisCoPatch-64.

Corruption	Corruption Intensity					Average
	1	2	3	4	5	
Brightness	91.4/37.2	91.7/34.0	92.7/29.5	93.9/24.4	94.8/21.4	92.9/29.3
Contrast	95.3/21.9	96.4/16.9	97.4/12.3	97.6/12.4	96.9/17.8	96.7/16.2
Defocus Blur	98.7/5.1	98.8/4.5	99.0/3.9	99.0/3.7	99.0/3.6	98.9/4.1
Elastic Transform	96.9/13.3	96.6/14.7	98.22/7.0	98.4/6.1	98.4/5.7	97.7/9.4
Fog	98.2/7.9	98.9/4.5	99.4/2.2	99.5/1.6	99.7/0.7	99.2/3.4
Frost	95.9/16.3	98.1/7.2	98.7/4.8	98.9/4.1	99.1/3.3	98.2/7.1
Gaussian Blur	98.3/7.0	98.8/4.4	99.0/3.8	99.0/3.6	99.1/3.5	98.8/4.5
Gaussian Noise	99.8/0.3	99.8/0.3	99.8/0.3	99.8/0.3	99.8/0.3	99.8/0.3
Glass Blur	98.9/4.2	99.2/2.8	99.5/1.6	99.5/1.4	99.5/1.4	99.3/2.3
Impulse Noise	99.8/0.5	99.8/0.3	99.8/0.3	99.8/0.3	99.8/0.3	99.8/0.4
JPEG Compression	83.6/55.4	83.1/54.0	82.9/53.4	81.2/53.6	78.1/56.9	81.8/54.7
Motion Blur	97.9/8.7	98.5/6.1	98.9/4.4	99.1/3.5	99.2/3.1	98.7/5.2
Pixelate	95.9/17.8	96.3/16.0	96.8/13.6	97.2/11.8	96.9/12.4	96.6/14.3
Saturate	95.8/18.6	98.5/6.2	93.7/29.5	96.4/16.0	97.4/11.1	96.3/16.3
Shot Noise	99.7/0.7	99.8/0.3	99.8/0.3	99.8/0.3	99.8/0.3	99.8/0.4
Snow	96.1/14.8	98.4/6.0	97.4/9.3	98.1/6.5	98.3/6.3	97.7/8.6
Spatter	93.0/31.3	95.1/19.3	97.3/9.9	94.8/17.7	96.6/11.2	95.3/17.9
Speckle Noise	99.5/2.0	99.6/1.4	99.6/1.6	99.5/1.7	99.6/1.6	99.5/1.6
Zoom Blur	98.4/6.7	98.6/5.3	98.8/4.5	98.9/4.1	99.0/3.7	98.8/4.9
Average	96.5/14.2	97.2/10.8	97.3/10.1	97.4/9.1	97.4/8.7	97.2/10.6