MOOD: Multi-level Out-of-distribution Detection Supplementary Material

Ziqian Lin*, Sreya Dutta Roy*, Yixuan Li
University of Wisconsin-Madison
Madison,Wisconsin
zlin284, duttaroy, sharonli@wisc.edu

A. Additional Information of Datasets

In Table 1 we provide additional information on the in-distribution and out-of-distribution datasets. We use the entire test splits for each of these datasets and provide the respective test set sizes. Each out-of-distribution input is preprocessed by subtracting the mean of in-distribution data and dividing the standard deviation. MNIST, fashion-MNIST and K-MNIST are padded by 2 in spatial dimensions and then extended to 3 channels. For STL10, SVHN, iSUN, Textures and Places365, the smaller sides of the images are resized to 32 and then center-cropped to 32×32. LSUN (crop) is a dataset created from LSUN by randomly cropping to 32×32 and LSUN (resize) is produced by downsampling each LSUN image to the size 32×32. As previously mentioned, these datasets span a range of complexities, and we present the average complexities per dataset. Table 1 also shows the negligible average compression latency per sample while the average inference time of a sample is 20 ms, using Nvidia 1080 Ti GPU card.

Dataset	# of Images	Mean Complexity (bytes)	Average Compression Latency (ms)
MNIST	10,000	456	0.426
K-MNIST	10,000	799	0.480
fashion-MNIST	10,000	917	0.448
LSUN (crop)	10,000	1,498	0.396
SVHN	10,000	1,736	0.339
Textures	5,640	2,165	0.348
STL10	8,000	2,222	0.338
CIFAR-100	10,000	2,247	0.352
Places365	328,500	2,255	0.348
CIFAR-10	10,000	2,271	0.355
iSUN	8,925	2,690	0.350
LSUN (resize)	10,000	2,695	0.346

Table 1. Additional information on the 12 datasets listed in the order of increasing complexity. The complexity is measured in bytes after PNG compression. The Compression Latency is measured on Intel(R) Core(TM) i7-7820X CPU @ 3.60GHz and the average inference time of each sample on Nvidia GPU 1080 Ti is 20 ms.

B. Experiment on JPEG2000

As seen previously, we choose the PNG compressor for encoding sample images and deriving bit lengths for optimal exit selection in the MOOD algorithm. In this section, we experiment with another lossless compressor JPEG2000. Figure 1 shows the complexity distribution of samples across the 12 datasets encoded using JPEG2000. The result of using JPEG2000 for MOOD is shown in Table 2. The JPEG2000 achieves competitive OOD detection results compared with PNG while using more inference time due to the lesser complexity distinguishability of JPEG2000.

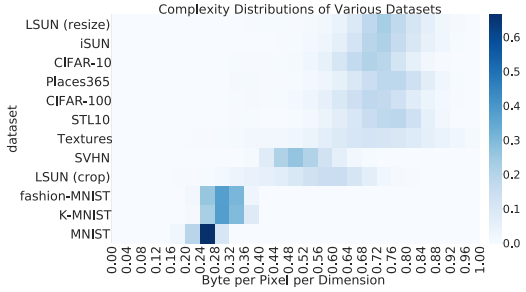

Figure 1. Complexity distribution using JPEG2000.

ID Dataset	Method	AUROC ↑	FLOPs ↓ ($\times 10^8$)
CIFAR-10	Exit@last	0.9048	1.05
	MOOD (PNG)	0.9129	0.79
	MOOD (JPEG2000)	0.9123	0.84
CIFAR-100	Exit@last	0.8451	1.05
	MOOD (PNG)	0.8507	0.79
	MOOD (JPEG2000)	0.8558	0.84

Table 2. OOD detection results of JPEG2000 compared to PNG and Exit@last.

C. Detailed Results for 10 OOD Datasets

In Table 3, we show detailed evaluation results for each of the 10 OOD datasets. We report performance of using both *constant* exiting at each exit, as well as the *dynamic* exit results with our MOOD algorithm.

OOD Dataset	ID Dataset	Measurement	Exit@1	Exit@2	Exit@3	Exit@4	Exit@5	MOOD
MNIST	CIFAR10	AUROC	0.9744	0.9875	0.9858	0.9889	0.9903	0.9979
		FPR95	0.1453	0.0546	0.0589	0.0542	0.0413	0.0036
	CIFAR100	AUROC	0.9059	0.9440	0.9589	0.9569	0.9451	0.9134
		FPR95	0.5505	0.2959	0.2491	0.2823	0.3103	0.5770
K-MNIST	CIFAR10	AUROC	0.9800	0.9839	0.9847	0.9868	0.9844	0.9986
		FPR95	0.0974	0.0805	0.0662	0.0586	0.0699	0.0033
	CIFAR100	AUROC	0.8654	0.9539	0.9410	0.9558	0.9416	0.9717
		FPR95	0.7756	0.2675	0.3616	0.2990	0.3676	0.1663
fashion-MNIST	CIFAR10	AUROC	0.9874	0.9876	0.9912	0.9930	0.9923	0.9991
		FPR95	0.0548	0.0504	0.0296	0.0219	0.0248	0.0011
	CIFAR100	AUROC	0.9705	0.9813	0.9810	0.9827	0.9795	0.9911
		FPR95	0.1524	0.0843	0.1061	0.1014	0.1226	0.0456
LSUN (crop)	CIFAR10	AUROC	0.9796	0.9821	0.9878	0.9877	0.9873	0.9923
		FPR95	0.0977	0.0953	0.0573	0.0609	0.0591	0.0320
	CIFAR100	AUROC	0.9439	0.9613	0.9610	0.9543	0.9495	0.9683
		FPR95	0.2709	0.2090	0.2176	0.2598	0.2784	0.1702
SVHN	CIFAR10	AUROC	0.8646	0.8990	0.9391	0.9497	0.9282	0.9649
		FPR95	0.7600	0.5554	0.4006	0.2892	0.3409	0.1716
	CIFAR100	AUROC	0.7418	0.8144	0.8364	0.8238	0.8126	0.8588
		FPR95	0.9077	0.8120	0.7778	0.7657	0.7756	0.6373
Textures	CIFAR10	AUROC	0.8060	0.8426	0.8732	0.8483	0.8233	0.8332
		FPR95	0.7259	0.6635	0.5856	0.6016	0.5512	0.5603
	CIFAR100	AUROC	0.6003	0.6408	0.6726	0.7073	0.7266	0.7169
		FPR95	0.9101	0.8780	0.8883	0.8851	0.8690	0.8683
STL10	CIFAR10	AUROC	0.6557	0.6733	0.6757	0.6422	0.6017	0.6131
		FPR95	0.8479	0.8324	0.8278	0.8438	0.8456	0.8439
	CIFAR100	AUROC	0.7185	0.7433	0.7588	0.7743	0.7744	0.7758
		FPR95	0.8538	0.8273	0.8150	0.8124	0.8131	0.7936
Places365	CIFAR10	AUROC	0.8923	0.9090	0.9128	0.8910	0.8609	0.8674
		FPR95	0.5004	0.4504	0.4216	0.4547	0.4568	0.4687
	CIFAR100	AUROC	0.7187	0.7622	0.7622	0.7656	0.7526	0.7567
		FPR95	0.8433	0.8014	0.8204	0.8283	0.8265	0.8237
iSUN	CIFAR10	AUROC	0.9282	0.9612	0.9476	0.9402	0.9384	0.9296
		FPR95	0.3978	0.2376	0.3190	0.3576	0.3179	0.3882
	CIFAR100	AUROC	0.6113	0.7304	0.7901	0.8068	0.7863	0.7784
		FPR95	0.9248	0.8069	0.7861	0.7394	0.7755	0.8147
LSUN (resize)	CIFAR10	AUROC	0.9409	0.9612	0.9468	0.9450	0.9412	0.9325
		FPR95	0.3433	0.2400	0.3362	0.3315	0.2911	0.3616
	CIFAR100	AUROC	0.6921	0.7816	0.8092	0.8035	0.7832	0.7760
		FPR95	0.8938	0.7365	0.7542	0.7384	0.7763	0.8122
Average	CIFAR10	AUROC	0.9009	0.9187	0.9245	0.9173	0.9048	0.9129
		FPR95	0.3970	0.3260	0.3103	0.3074	0.2999	0.2834
	CIFAR100	AUROC	0.7769	0.8313	0.8471	0.8531	0.8451	0.8507
		FPR95	0.7083	0.5719	0.5776	0.5712	0.5915	0.5709

Table 3. Results for 10 OOD datasets. Metrics are AUROC and FPR@95.