INTERPRETABLE OUT-OF-DISTRIBUTION DETECTION USING PATTERN DETECTION
- SUPPLEMENTARY MATERIAL -

Anonymous authors
Paper under double-blind review

Table 1 and 2 summarize all the experiments that have been carried out in this work. In particular, we also used the CIFAR100 Krizhevsky (2009), with images upsampled to 224×224 during our cross-dataset and perturbation experiments, and tested our Particul-based confidence measures with 2 detectors.

1 EXPERIMENTS ON CIFAR100

In addition to what has been stated in the main paper regarding the stability of C_M^{fssd}, this method gives opposite results for Caltech101 v. CIFAR100 (Table 1) depending on the choice of validation OoD dataset: for $D_{ood,val} =$ CUB200, C_M^{fssd} is a perfect classifier (AUROC=100) between Caltech101 and CIFAR100 inputs, *i.e.,*

$$\forall (x_{iod}, x_{ood}) \in D_{iod} \times D_{ood},\ C_M(x_{iod}) > C_M(x_{ood})$$

However, for $D_{ood,val} =$ Stanford Cars, the distributions are reversed (AUROC=0), *i.e.,*

$$\forall (x_{iod}, x_{ood}) \in D_{iod} \times D_{ood},\ C_M(x_{iod}) < C_M(x_{ood})$$

Note also that for D_{iod}=CIFAR100, C_M^{fnrd} is a perfect distinguisher *w.r.t.* to other datasets. For all other experiments (cross-dataset OoD with other measures, perturbation OoD detection), the results on CIFAR100 are consistent with the other datasets:

1. C_M^{cP} performs better than C_M^{vP} on cross-dataset detection;
2. C_M^{fnrd} is inversely correlated to blur and brightness changes;
3. C_M^{MCP}/C_M^{EB} are not correlated to noise.

2 INFLUENCE OF THE NUMBER OF DETECTORS

Training 2, 4 or 6 detectors, either globally or per-class, seems to have a small influence on the robustness of our Particul-based confidence measure. Indeed, the additional information (confidence score) provided by an additional detector is usually counter-balanced by the difficulty of learning a new pattern distinct from all the other detectors. This indicates that increasing the number of detectors beyond 6 may become counter-productive during training.

REFERENCES

Alex Krizhevsky. Learning multiple layers of features from tiny images. Technical report, 2009. 1

			OoD-specific				OoD-agnostic								
D_{iod}	D_{ood}	Metrics	FSSD ($D_{ood,val}$)				MCP	EB	fNRD	*vanilla* Particul			*class-based* Particul		
			CT	CB	SC	CF				$P = 2$	$P = 4$	$P = 6$	$P = 2$	$P = 4$	$P = 6$
CT	CB	AUROC↑	/	(95.0)	28.6	92.5	**87.2**	83.0	46.8	71.5 ± 3.9	72.5 ± 4.8	71.7 ± 1.4	74.4 ± 1.9	72.7 ± 0.8	69.7 ± 1.7
		AUPR↑	/	(88.5)	12.1	81.6	**77.5**	70.5	17.7	43.8 ± 4.8	44.0 ± 8.1	47.3 ± 4.6	41.5 ± 1.8	38.6 ± 2.3	33.9 ± 2.4
		FPR80↓	/	(04.5)	94.5	09.8	**22.1**	34.1	86.3	54.8 ± 5.1	49.9 ± 7.6	57.3 ± 1.8	47.4 ± 3.4	49.5 ± 1.1	54.6 ± 2.2
	SC	AUROC↑	/	41.3	(78.0)	46.4	**85.3**	84.0	70.1	52.2 ± 10.7	47.3 ± 10.5	41.0 ± 2.4	63.4 ± 3.1	67.1 ± 0.6	67.0 ± 2.8
		AUPR↑	/	15.3	(49.2)	13.5	**71.8**	67.1	44.8	23.6 ± 8.3	19.9 ± 8.6	16.6 ± 3.5	20.6 ± 2.2	24.0 ± 1.0	24.4 ± 2.3
		FPR80↓	/	96.4	(45.2)	83.9	**26.9**	32.5	68.1	86.4 ± 12.4	88.9 ± 6.0	95.5 ± 1.5	62.9 ± 4.1	58.1 ± 0.8	60.5 ± 4.2
	CF	AUROC↑	/	100	0.0	(100)	87.1	71.7	11.9	85.7 ± 5.5	93.1 ± 2.3	**96.1** ±1.3	91.3 ± 0.2	92.1 ± 2.2	91.0 ± 2.0
		AUPR↑	/	99.9	6.0	(99.7)	74.2	54.7	6.4	56.1 ± 13.5	73.2 ± 8.1	**82.7** ±4.4	67.6 ± 0.7	70.3 ± 8.3	66.7 ± 4.6
		FPR80↓	/	0.0	100	(0.0)	22.2	72.6	97.1	27.5 ± 10.6	10.5 ± 5.1	**5.5** ±2.4	14.4 ± 1.0	11.9 ± 4.6	14.9 ± 4.3
CF	CT	AUROC↑	(99.4)	99.2	99.3	/	69.1	67.0	**100**	68.6 ± 0.3	62.2 ± 1.1	59.2 ± 2.1	68.1 ± 0.7	66.6 ± 0.8	66.4 ± 1.0
		AUPR↑	(99.9)	99.7	99.9	/	94.6	94.4	**100**	94.4 ± 0.1	92.7 ± 0.1	91.8 ± 0.6	93.9 ± 0.4	93.5 ± 0.2	93.3 ± 0.2
		FPR80↓	(0.6)	0.7	0.5	/	62.7	69.0	**0**	60.6 ± 0.4	69.5 ± 2.2	73.1 ± 2.2	57.0 ± 1.5	59.2 ± 1.9	59.2 ± 2.7
	CB	AUROC↑	96.0	(96.6)	92.9	/	70.1	69.5	**99.9**	68.9 ± 2.1	67.5 ± 6.0	62.6 ± 4.4	72.3 ± 2.3	70.2 ± 0.8	67.2 ± 1.2
		AUPR↑	97.4	(97.8)	95.1	/	81.8	81.2	**100**	79.5 ± 1.6	77.8 ± 4.7	72.1 ± 3.8	81.1 ± 1.7	80.1 ± 0.8	77.0 ± 0.9
		FPR80↓	5.9	(4.6)	11.3	/	61.6	61.9	**0**	58.3 ± 3.2	58.6 ± 9.7	64.4 ± 3.8	48.7 ± 2.6	56.1 ± 2.6	58.8 ± 3.0
	SC	AUROC↑	99.9	99.6	(99.9)	/	65.5	66.0	**100**	54.2 ± 0.7	55.9 ± 2.6	49.5 ± 6.0	73.4 ± 2.6	78.6 ± 3.0	77.9 ± 2.6
		AUPR↑	99.9	99.7	(99.9)	/	72.4	73.7	**100**	65.1 ± 1.1	65.2 ± 2.2	58.1 ± 5.2	79.3 ± 2.0	82.6 ± 2.9	82.2 ± 1.7
		FPR80↓	0	0.1	(0)	/	68.8	67.9	**0**	85.6 ± 1.3	81.4 ± 2.8	86.1 ± 5.5	52.6 ± 5.6	40.2 ± 6.2	41.9 ± 5.8
CB	CT	AUROC↑	(94.7)	/	94.5	78.5	**96.1**	94.0	72.1	94.2 ± 1.8	91.8 ± 3.3	93.7 ± 0.4	88.7 ± 4.2	93.6 ± 0.3	93.1 ± 2.0
		AUPR↑	(98.3)	/	98.3	92.8	**99.1**	98.2	91.5	98.1 ± 0.7	97.5 ± 0.8	97.7 ± 0.2	96.4 ± 1.3	98.3 ± 0.1	98.1 ± 0.6
		FPR80↓	(7.0)	/	6.4	35.9	**5.1**	8.5	52.8	9.1 ± 3.3	12.9 ± 5.7	9.5 ± 1.3	19.2 ± 8.4	9.4 ± 1.6	12.8 ± 5.0
	SC	AUROC↑	98.6	/	(99.3)	84.7	98.8	97.1	88.5	98.7 ± 0.7	97.9 ± 1.8	**99.1** ±0.1	92.7 ± 6.0	94.4 ± 3.9	97.5 ± 1.6
		AUPR↑	98.6	/	(98.9)	84.1	**98.6**	96.9	88.8	98.0 ± 1.0	97.3 ± 1.9	**98.6** ±0.1	90.5 ± 6.3	93.5 ± 4.6	96.0 ± 2.0
		FPR80↓	0.4	/	(0.5)	25.8	**0.4**	1.5	18.0	1.4 ± 0.6	1.9 ± 1.5	0.9 ± 0.1	13.3 ± 13.2	7.9 ± 8.5	4.9 ± 4.6
	CF	AUROC↑	68.0	/	80.9	(97.6)	**98.0**	82.8	71.0	93.4 ± 1.9	92.3 ± 0.7	94.2 ± 1.7	90.8 ± 4.9	94.3 ± 1.0	96.5 ± 0.1
		AUPR↑	48.7	/	66.9	(96.1)	**97.4**	70.9	68.2	88.8 ± 3.4	86.1 ± 0.6	88.5 ± 3.3	83.1 ± 9.3	91.2 ± 1.7	94.7 ± 0.3
		FPR80↓	52.1	/	31.6	(2.4)	**1.0**	28.2	68.7	10.9 ± 3.1	12.2 ± 1.0	8.9 ± 2.9	15.0 ± 9.0	9.2 ± 1.6	4.3 ± 0.8
SC	CT	AUROC↑	(99.9)	99.1	/	98.1	96.7	**98.9**	45.5	97.5 ± 1.1	97.4 ± 1.9	98.4 ± 0.6	85.2 ± 12.4	86.8 ± 1.6	94.0 ± 3.3
		AUPR↑	(100)	99.8	/	99.6	99.4	**99.7**	83.1	99.5 ± 0.2	99.5 ± 0.4	**99.7** ±0.1	97.0 ± 2.5	97.5 ± 0.4	98.9 ± 0.7
		FPR80↓	(0.1)	0.8	/	2.0	1.9	**0.6**	79.0	3.9 ± 2.2	3.9 ± 2.9	1.8 ± 0.7	29.1 ± 25.6	21.0 ± 6.8	8.0 ± 5.5
	CB	AUROC↑	100	(99.9)	/	99.9	97.6	99.6	51.8	99.3 ± 0.5	99.6 ± 0.5	**99.8** ±0.2	79.4 ± 21.9	95.4 ± 0.8	97.0 ± 0.9
		AUPR↑	100	(99.9)	/	99.9	98.0	99.7	56.6	99.5 ± 0.3	99.7 ± 0.3	**99.9** ±0.1	91.4 ± 4.2	96.7 ± 0.7	97.9 ± 0.7
		FPR80↓	0	(0)	/	0	3.7	0.4	72.5	0.4 ± 0.4	0.1 ± 0.1	**0.0** ±0.1	20.2 ± 9.6	7.2 ± 2.9	2.7 ± 0.9
	CF	AUROC↑	100	100	/	(100)	92.9	93.9	39.0	**100** ±0.0	**100**±0.1	**100** ±0.0	87.5 ± 11.7	85.9 ± 14.2	93.5 ± 5.0
		AUPR↑	100	100	/	(100)	94.0	94.3	39.3	**100** ±0.0	99.9 ± 0.1	**100** ±0.0	86.2 ± 12.4	85.7 ± 14.2	94.1 ± 4.1
		FPR80↓	0	0	/	(0)	3.9	5.4	87.6	**0.0** ±0.0	**0.0** ±0.0	**0.0** ±0.0	24.2 ± 20.3	19.9 ± 26.3	10.6 ± 9.4

Table 1: AUROC↑, AUPR↑ and FPR80↓ scores for different pairs (D_{iod}, D_{ood}). For readability, all scores are given on a 0-100 scale. SC denotes StanfordCars, CB denotes CUB200, CT denotes Caltech101 and CF denotes CIFAR100. The thick vertical line separates FSSD (OoD-specific) from all other agnostic methods. The double horizontal line separates experiments where D_{iod} is heterogenous (Caltech101 and CIFAR100, top) or homogeneous (CUB200 and Standford Cars, bottom). For FSSD scores, we indicate the validation dataset $D_{ood,val}$ used during the calibration, and scores in parenthesis correspond to the optimal case $D_{ood,val} \sim \mathcal{D}_{ood}$. In red, we highlight the contradictory results obtained when using two different validation datasets during the linear regression phase of the FSSD calibration. For Particul-based measures, we provide the scores averaged over 3 random initializations of the detectors along with corresponding unbiased standard deviation. For OoD-agnostic approaches, we indicate in bold the best performing measure for each experiment and each metric. Best viewed in color.

Pert. P	Arch.	$\mathcal{D}_{iod}$	MCP	EB	fNRD	*vanilla* Particul			*class-based* Particul		
						$P=2$	$P=4$	$P=6$	$P=2$	$P=4$	$P=6$
Blur (-)	CT	Resnet50	-1.00	-0.85	1.00	-0.18 ± 1.02	-0.91 ± 0.06	-0.95 ± 0.00	-0.96 ± 0.03	-0.95 ± 0.00	-0.95 ± 0.00
	CF	Resnet50	-0.95	-0.95	1.00	-0.98 ± 0.03	-1.00 ± 0.00	-1.00 ± 0.00	-0.95 ± 0.00	-0.95 ± 0.00	-0.95 ± 0.00
	CB	Resnet50	-0.90	0.61	1.00	0.95 ± 0.00	0.95 ± 0.00	0.95 ± 0.00	0.96 ± 0.03	0.95 ± 0.00	1.00 ± 0.00
		VGG19	-0.86	-1.00	1.00	-0.95 ± 0.05	-1.00 ± 0.00	-1.00 ± 0.00	-0.91 ± 0.03	-1.00 ± 0.00	-0.96 ± 0.06
	SC	Resnet50	-1.00	-0.95	1.00	-0.86 ± 0.03	-0.25 ± 0.60	0.50 ± 0.14	-0.85 ± 0.00	-0.85 ± 0.00	-0.86 ± 0.03
Noise (-)	CT	Resnet50	0.05	0.50	-1.00	-1.00 ± 0.00	-1.00 ± 0.00	-1.00 ± 0.00	-0.98 ± 0.01	-1.00 ± 0.00	-1.00 ± 0.00
	CF	Resnet50	-0.07	0.64	-1.00	-0.83 ± 0.05	-0.96 ± 0.03	-0.98 ± 0.00	-0.87 ± 0.10	-0.90 ± 0.06	-0.96 ± 0.05
	CB	Resnet50	-0.24	0.12	-1.00	-0.66 ± 0.05	-0.69 ± 0.00	-0.68 ± 0.08	-0.73 ± 0.10	-0.78 ± 0.10	-0.67 ± 0.08
		VGG19	-1.00	-1.00	-1.00	-0.98 ± 0.01	-1.00 ± 0.00	-1.00 ± 0.00	-0.90 ± 0.12	-0.92 ± 0.06	-0.97 ± 0.05
	SC	Resnet50	-0.29	-0.29	-1.00	-0.75 ± 0.07	-0.77 ± 0.03	-0.82 ± 0.01	-0.93 ± 0.04	-0.61 ± 0.61	-0.93 ± 0.12
Brightness (+)	CT	Resnet50	0.98	1.00	-1.00	1.00 ± 0.00	0.97 ± 0.04	0.98 ± 0.04	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00
	CF	Resnet50	0.98	1.00	-1.00	0.96 ± 0.03	0.99 ± 0.01	0.97 ± 0.04	0.98 ± 0.01	0.98 ± 0.00	0.98 ± 0.00
	CB	Resnet50	1.00	1.00	-0.31	-0.20 ± 0.06	-0.01 ± 0.52	0.23 ± 0.11	0.94 ± 0.05	0.78 ± 0.20	0.67 ± 0.20
		VGG19	0.98	1.00	-0.98	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00
	SC	Resnet50	1.00	1.00	-1.00	1.00 ± 0.00	0.98 ± 0.03	0.90 ± 0.18	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00
Rotation forth (-)	CT	Resnet50	-0.30	-0.28	-0.24	-0.46 ± 0.01	-0.41 ± 0.18	-0.42 ± 0.05	-0.26 ± 0.20	-0.51 ± 0.08	-0.52 ± 0.07
	CF	Resnet50	-0.76	-0.37	-0.19	-0.79 ± 0.05	-0.79 ± 0.07	-0.53 ± 0.14	-0.75 ± 0.03	-0.81 ± 0.06	-0.74 ± 0.04
	CB	Resnet50	-0.95	-0.89	-0.42	-0.94 ± 0.00	-0.82 ± 0.20	-0.94 ± 0.01	-0.92 ± 0.01	-0.94 ± 0.01	-0.93 ± 0.00
		VGG19	-0.95	-0.94	-0.42	-0.95 ± 0.00	-0.95 ± 0.00	-0.95 ± 0.00	-0.95 ± 0.00	-0.95 ± 0.00	-0.95 ± 0.00
	SC	Resnet50	-0.76	-0.76	-0.56	-0.70 ± 0.02	-0.71 ± 0.06	-0.67 ± 0.09	-0.84 ± 0.13	-0.86 ± 0.08	-0.89 ± 0.08
Rotation back (+)	CT	Resnet50	0.52	0.65	0.20	0.41 ± 0.07	0.51 ± 0.09	0.44 ± 0.09	0.67 ± 0.10	0.68 ± 0.10	0.73 ± 0.06
	CF	Resnet50	0.75	0.38	-0.02	0.79 ± 0.11	0.72 ± 0.15	0.62 ± 0.35	0.91 ± 0.02	0.88 ± 0.02	0.71 ± 0.08
	CB	Resnet50	0.95	0.90	0.32	0.95 ± 0.00	0.77 ± 0.32	0.94 ± 0.01	0.98 ± 0.03	0.98 ± 0.03	0.98 ± 0.03
		VGG19	0.95	0.93	0.35	0.94 ± 0.01	0.94 ± 0.01	0.92 ± 0.01	0.94 ± 0.01	0.92 ± 0.01	0.92 ± 0.01
	SC	Resnet50	0.72	0.72	0.32	0.69 ± 0.05	0.78 ± 0.06	0.74 ± 0.03	0.84 ± 0.03	0.84 ± 0.05	0.89 ± 0.07

Table 2: Spearman rank correlation coefficient between the intensity λ of a perturbation applied to dataset $\mathcal{D}$ and the average confidence measure $\Gamma(C_M, \mathcal{D}, P, \lambda)$ computed over the dataset. For each perturbation P, the expected correlation sign is recalled in parenthesis. In red, we highlight experiments where the distributions are either uncorrelated (small correlation coefficient), inversely correlated (opposite coefficient sign) or unstable (high standard deviation for Particul-based measures). Best viewed in color.