Out-of-Distribution Detection Using an Ensemble of Self Supervised Leave-out Classifiers

Apoorv Vyas[1,3]*, Nataraj Jammalamadaka[1]* Xia Zhu[2]*, Dipankar Das[1], Bharat Kaul[1], Theodore L. Willke[2]

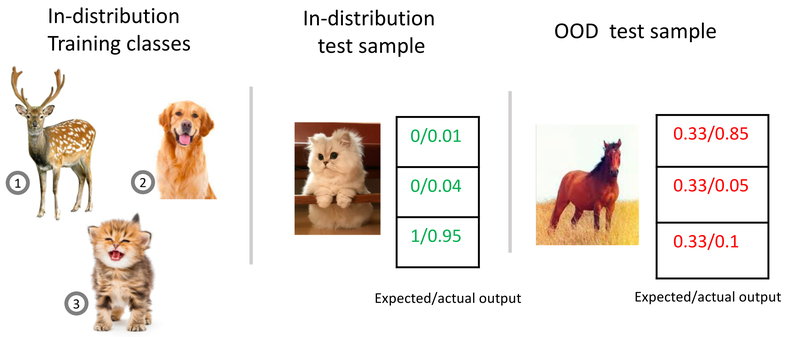
[1] Intel labs, Bangalore, India, [2] Intel labs, Hillsboro, OR 97124, USA, [3] Idiap Research Institute, Switzerland

* indicates equal contribution. Work done when the authors were working at Intel labs

ECCV 2018 European Conference on Computer Vision

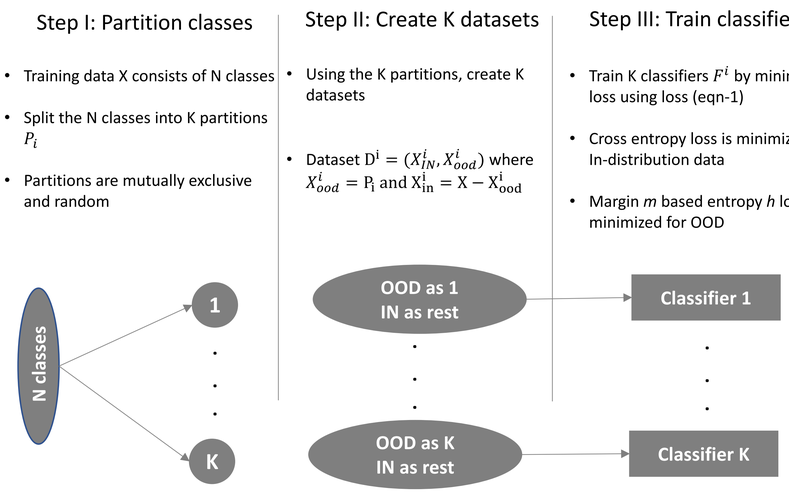
8 – 14 September 2018 | Munich, Germany

1 Motivation



- **Out-of-distribution data (OOD):** Any input that is sampled from a distribution that the network has not been trained on.
- **Challenge:** Training data has no access to OOD samples and further, #OOD classes >> #In classes.
- **Limitation:** Deep neural networks overestimate the class confidence scores.
- **Prior work:** Softmax baseline, Uncertainty estimation, Rescaling softmax scores.

2 Training method



$$\mathcal{L} = -\frac{1}{|X_{in}|}\sum_{x_i \in X_{in}} log\left(F_{y_i}(x_i)\right) + \beta * max\left(m + \frac{\sum_{x_i \in X_{in}} H(F(x_i))}{|X_{in}|} - \frac{\sum_{x_o \in X_{ood}} H(F(x_o))}{|X_{ood}|}, 0\right)$$

cross-entropy | margin-entropy

3 Testing method

Algorithm 2: Algorithm for OOD Detection using K Leave Out Classifiers

Input : Test Image x_t, K leave-out Classifiers $F_i, i \in 1, ..., K$, temperature scaled versions $F_i(x_t; T)$, perturbation factor ϵ, number of classes N

Output : Classification score vector S_t, OOD score O_t

$S_t = \sum_K F_i(x_t)$

OOD Score $= \sum_K \max_N (F_i(\hat{x}^i, T)) - H(F_i(\hat{x}_i; T))$

where $\hat{x_t}^i = x_t^i - \epsilon * sign(\frac{\partial (H(F_i(\hat{x}_i; T)))}{\partial x})$

4 Results

4.1 Ablation studies

Loss	SFX/84.1	SFX+H/50.7	SFX+Margin H/23.0	
Splits	3/32.4	5/23.0	10/28.7	20/23.9
Detection score	SFX/50.5	H/36.2	SFX+H/ 38.6	(SFX+H)@T/23.0

- Softmax + Entropy loss with Margin improves both classification accuracy as well as OOD detection over other losses.
- At 3 splits, our method outperforms SOTA on both cifar-10 and cifar-100.

4.2 Main Results

	OOD Dataset	FPR at 95% TPR ↓	Detection Error ↓	AUPR In ↑	AUPR Out ↑
		each cell in **ODIN**/**Our Method** format			
DSN-BC CIFAR-10	TINc	4.3/**1.2**	4.7/**2.6**	99.1/**99.7**	99.1/**99.6**
	TINr	7.5/**2.9**	6.1/**3.9**	98.6/**99.4**	98.5/**99.3**
	LSUNc	8.7/**3.4**	6.0/**4.1**	98.5/**99.3**	97.8/**99.3**
	LSUNr	3.8/**0.8**	4.4/**2.1**	99.3/**99.8**	99.2/**99.7**
	GSSN	**0.0**/**0.0**	**0.5**/0.2	**100**/99.9	**99.9**/99.6
DSN-BC CIFAR-100	TINc	17.3/**8.3**	8.8/**6.3**	97.4/**98.6**	96.8/**98.3**
	TINr	44.3/**20.5**	17.5/**10.0**	91.4/**96.7**	90.1/**95.8**
	LSUNc	17.6/**14.7**	9.4/**8.5**	97.1/**97.6**	96.5/**97.2**
	LSUNr	44.0/**16.2**	16.8/**8.8**	92.4/**97.4**	90.6/**96.6**
	GSSN	**0.2**/38.5	**1.9**/8.2	**99.7**/96.4	**99.1**/90.0
WRN-28-10 CIFAR-10	TINc	23.4/**0.8**	11.6/**2.2**	92.8/**99.8**	94.7/**99.8**
	TINr	25.5/**2.9**	13.4/**3.8**	89.0/**99.4**	93.6/**99.4**
	LSUNc	21.8/**1.9**	9.8/**3.2**	95.8/**99.6**	95.5/**99.6**
	LSUNr	17.6/**0.9**	9.7/**2.5**	93.8/**99.7**	96.1/**99.7**
	GSSN	**0.0**/**0.0**	**0.10**/1.0	**100**/99.7	**100**/99.2
WRN-28-10 CIFAR-100	TINc	43.9/**9.2**	17.2/**6.7**	91.4/**98.4**	90.0/**98.1**
	TINr	55.9/**24.5**	23.3/**11.6**	82.8/**95.5**	84.4/**94.8**
	LSUNc	39.6/**14.2**	15.6/**8.2**	92.4/**97.6**	91.6/**97.7**
	LSUNr	56.5/**16.5**	21.7/**9.1**	86.2/**97.0**	84.9/**96.4**
	GSSN	**1.0**/98.3	**2.9**/16.9	**99.1**/88.6	**95.9**/71.6

Table 1: Distinguishing in- and out-of-distribution test set data for the image classification. All values are percentages.

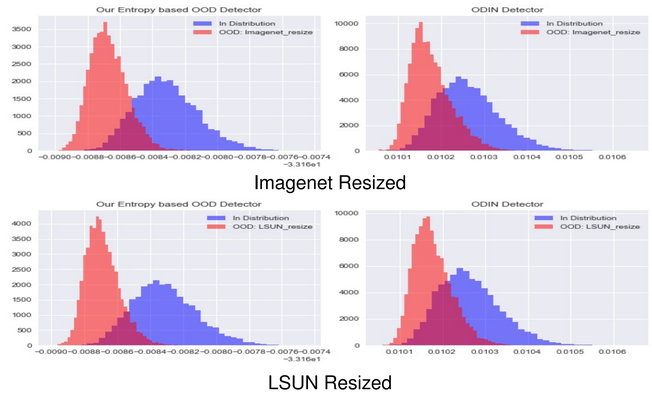


Figure 1: Histogram of ID and OOD detection scores with ours v.s. ODIN detector:

- **Evaluation metrics** - FPR@TPR, AUPR In, AUPR Out, Detection error (Min. misclassification rate over thresholds).
- Our method significantly outperforms ODIN (Table-1) on all metrics, networks, and most of the datasets.
- Improvements are better on more complex datasets like resized Tiny ImageNet and LSUN.
- Using a single dataset (iSUN) as validation set, our method is able to generalize to various other OOD datasets.

4.3 Mean and standard deviation of FPR at 95% TPR

OOD Dataset	DenseNet CIFAR-10	DenseNet CIFAR-100	WRN-28-10 CIFAR-10	WRN-28-10 CIFAR-100
Tiny-Imagenet Crop	1.5 ± 0.2	10.3 ± 1.3	1.3 ± 0.3	10.3 ± 2.2
Tiny-Imagenet resize	3.9 ± 0.8	26.6 ± 4.2	4.6 ± 1.3	29.8 ± 5.1
LSUN crop	4.5 ± 1.4	16.9 ± 1.3	3.8 ± 1.2	15.5 ± 1.4
LSUN resize	1.3 ± 0.6	20.2 ± 2.8	1.5 ± 0.4	22.5 ± 6.1
Gaussian noise	27.1 ± 40.0	82.2 ± 12.8	31.5 ± 33.9	67.5 ± 44.3

- Our method is not very sensitive to random class division and convincingly outperforms on SOTA.
- In practice, several class partitions can be tried and the best model can be selected based on validation data.

5 Conclusions and Future Work

- Softmax + Entropy loss with Margin improves both classification accuracy as well as OOD detection over other losses.
- At low number of partitions (as low as 3), the proposed method outperforms SOTA.
- The proposed method requires large memory and computational resources. To alleviate this, we will explore parameter sharing and then add branches for individual classifiers.