Reducing Network Agnostophobia

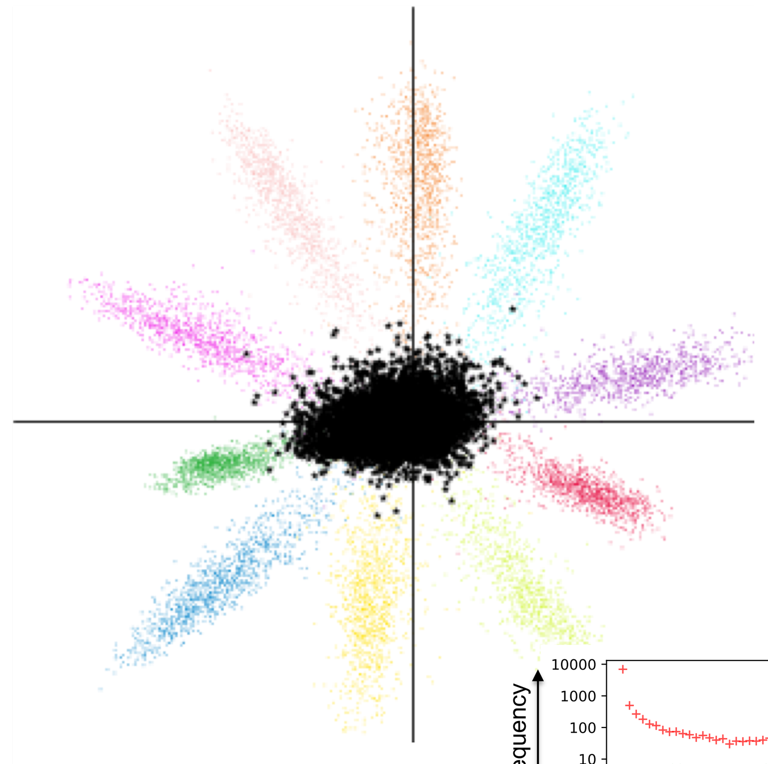
Akshay Raj Dhamija, Manuel Günther & Terrance E. Boult
{ adhamija, mgunther, tboult } @ vast.uccs.edu
UCCS Vision and Security Technology
UNIVERSITY OF COLORADO COLORADO SPRINGS

Abstract

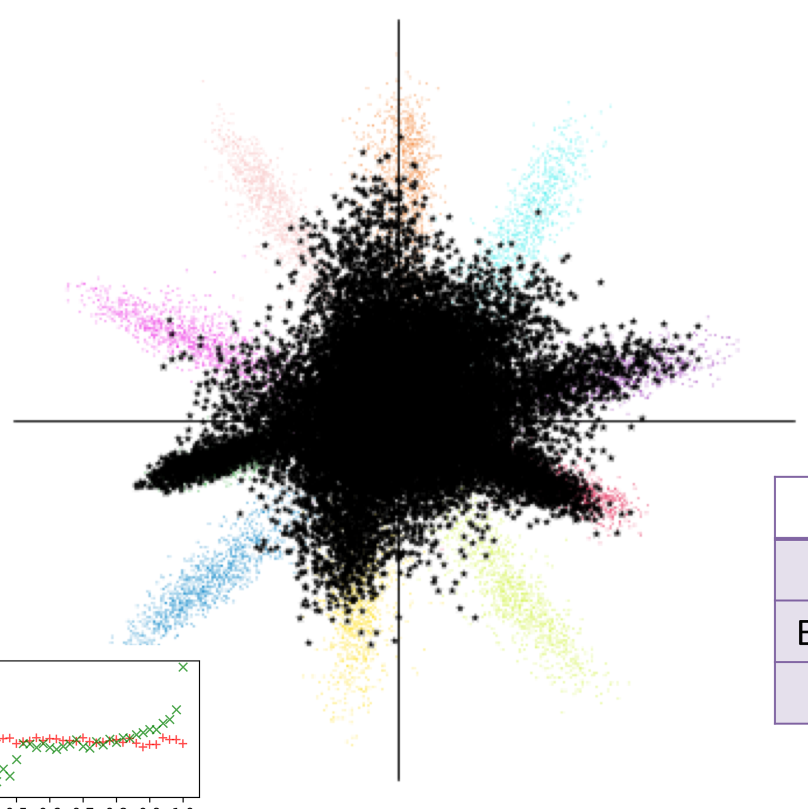
***Agnostophobia**, the fear of the unknown, can be experienced by deep learning engineers while applying their networks to real-world applications. Unfortunately, network behavior is not well defined for inputs far from a networks training set. In an uncontrolled environment, networks face many instances that are not of interest to them and have to be rejected in order to avoid a false positive. This problem has previously been tackled by researchers by either a) thresholding softmax, which by construction cannot return "none of the known classes", or b) using an additional background or garbage class. In this paper, we show that both of these approaches help, but are generally insufficient when previously unseen classes are encountered. We also introduce a new evaluation metric that focuses on comparing the performance of multiple approaches in scenarios where such unseen classes or unknowns are encountered. Our major contributions are simple yet effective Entropic Open-Set and Objectosphere losses that train networks using negative samples from some classes. These novel losses are designed to maximize entropy for unknown inputs while increasing separation in deep feature space by modifying magnitudes of known and unknown samples. Experiments on networks trained to classify classes from MNIST and CIFAR-10 show that our novel loss functions are significantly better at dealing with unknown inputs from datasets such as Devanagari, NotMNIST, CIFAR-100, and SVHN.*

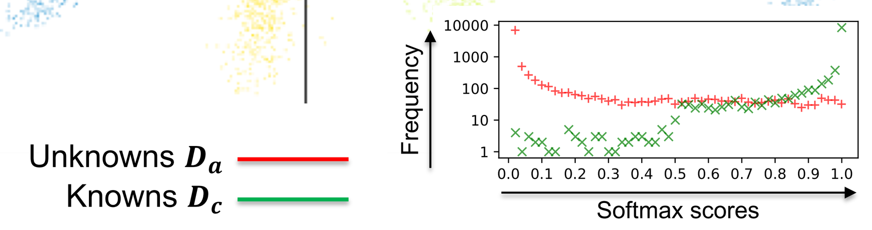
Default Behavior of Deep Networks

Easy Samples CIFAR

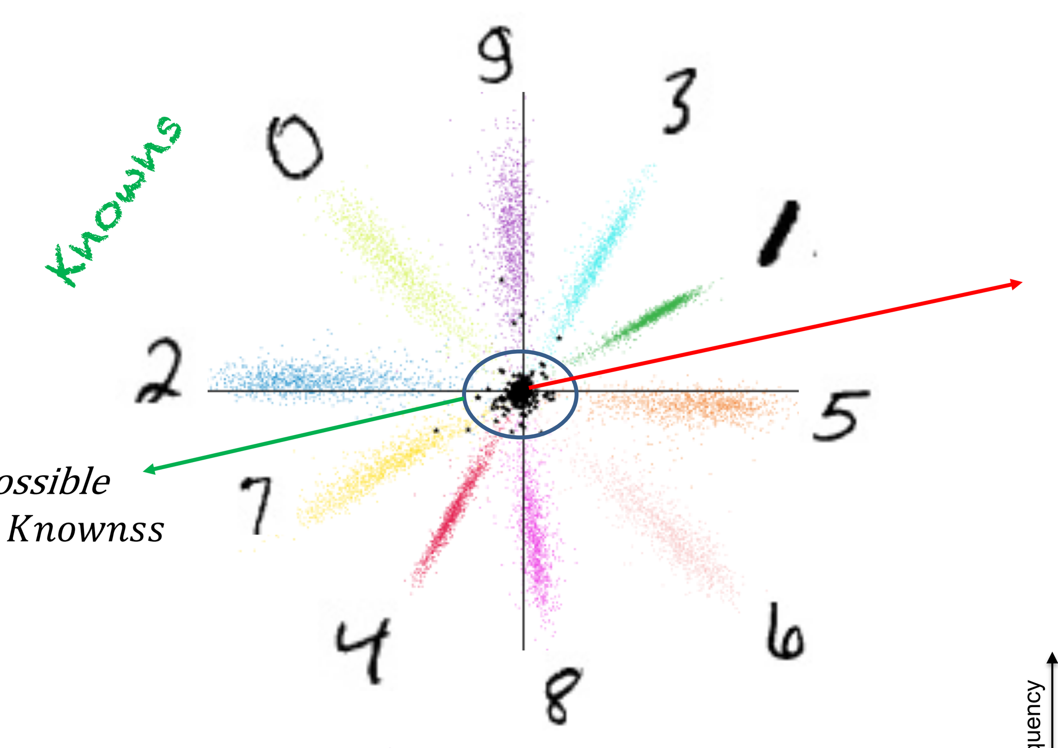


Hard Samples NIST Letters

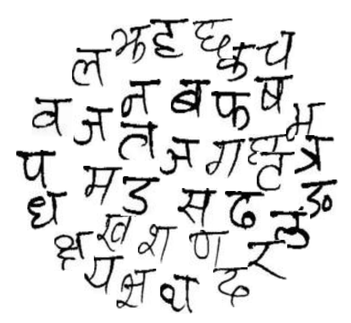




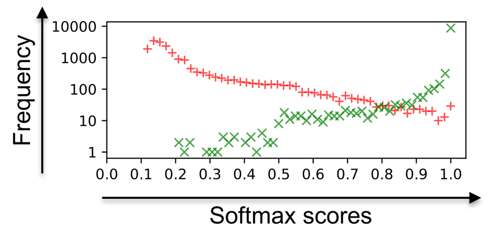
Our Approach



Unknowns



Features of Unknown samples pushed to the center



Entropic Open-Set Loss

Increases entropy of the softmax scores for unknwons

$$\mathcal{J}_E(x) = \begin{cases} -\log S_c(x) & \text{if } x \in \mathcal{D}'_c \text{ and } x \text{ is from class } c \\ -\frac{1}{c}\sum_{i=1}^{c} \log S_i(x) & \text{if } x \in \mathcal{D}'_b \end{cases}$$

Objectosphere Loss

Minimizes euclidean length of deep representations for unknowns

$$\mathcal{J}_R(x) = \mathcal{J}_E + \begin{cases} \max(\xi - \| F(x) \|, 0)^2 & \text{if } x \in \mathcal{D}'_c \\ \| F(x) \|^2 & \text{if } x \in \mathcal{D}'_b \end{cases}$$

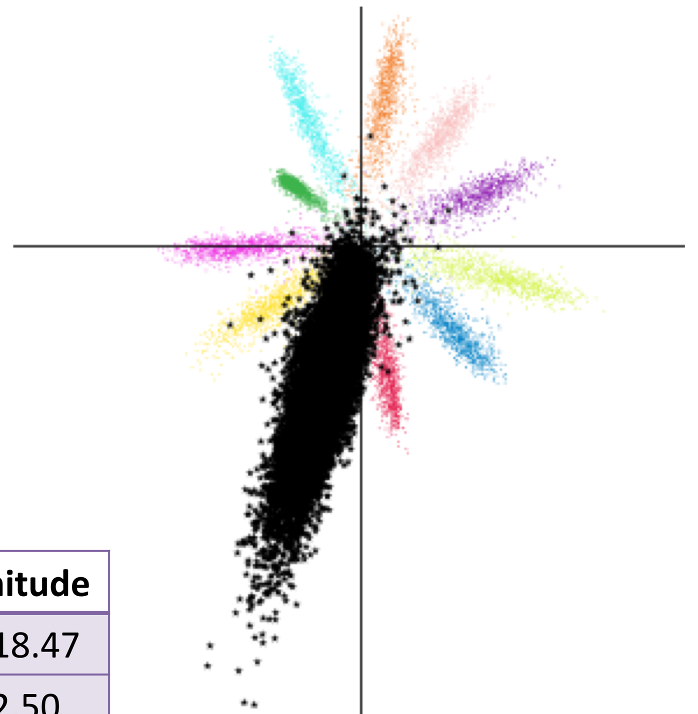
Observations

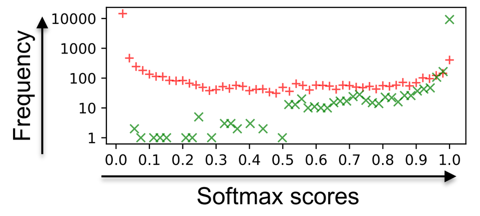
Magnitude of Deep Feature Representation of
Known Samples > Unknown Samples

Entropy of
Known Samples < Unknown Samples

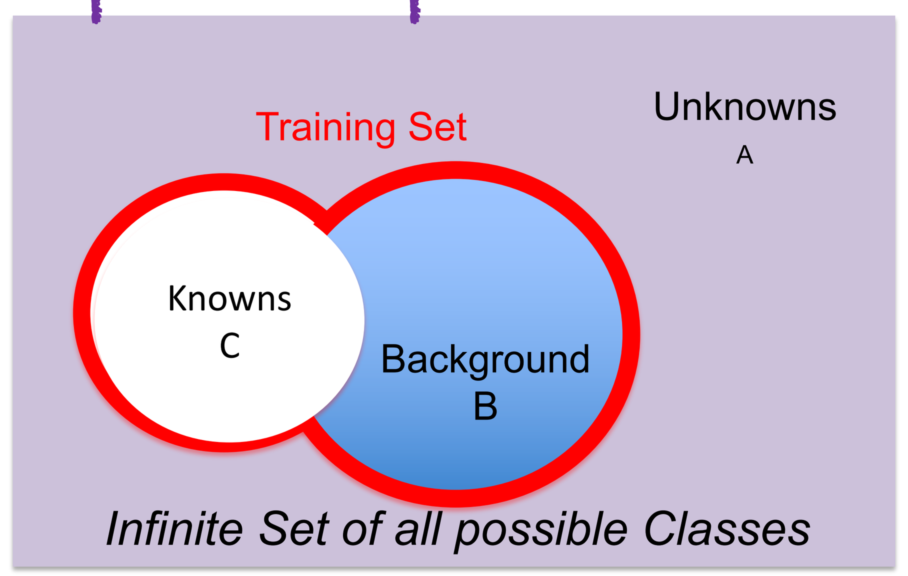
Algorithm	D_c Entropy	D_a Entropy	D_c Magnitude	D_a Magnitude
Softmax	0.015 ± .084	0.318 ± .312	94.90 ± 27.47	32.27 ± 18.47
Entropic Open-Set	0.050 ± .159	1.984 ± .394	50.14 ± 17.36	1.50 ± 2.50
Objectosphere	0.056 ± .168	2.031 ± .432	76.80 ± 28.55	2.19 ± 4.73

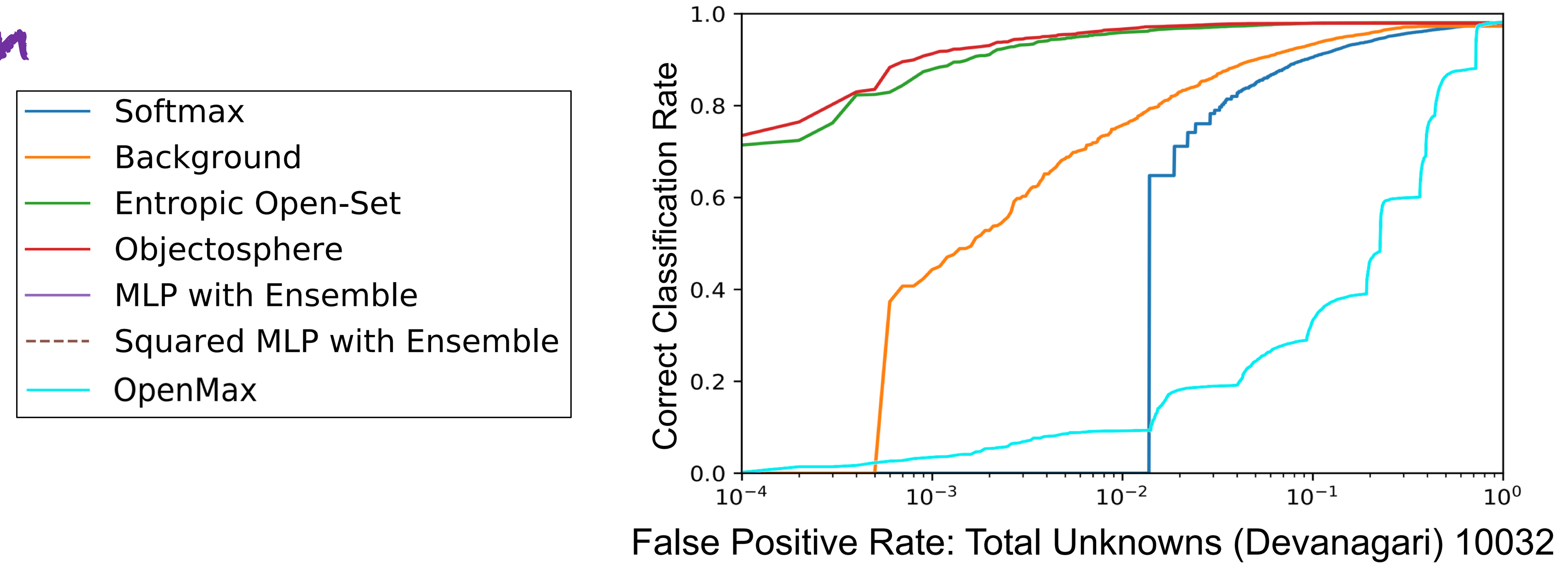
Background Class

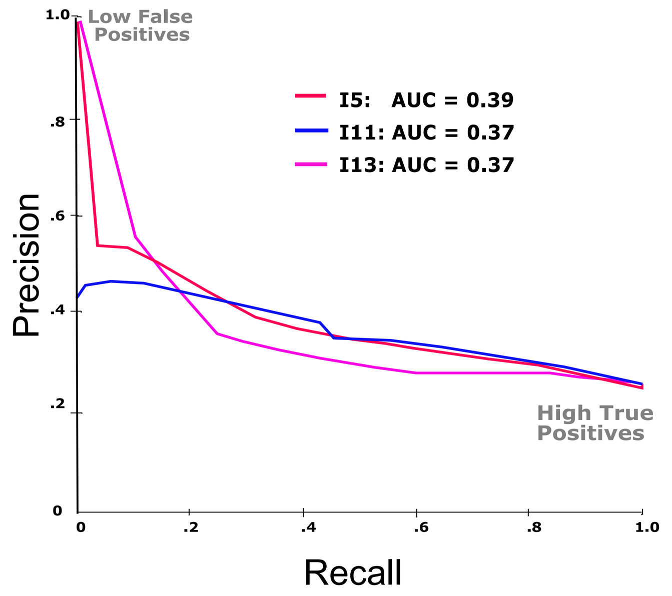


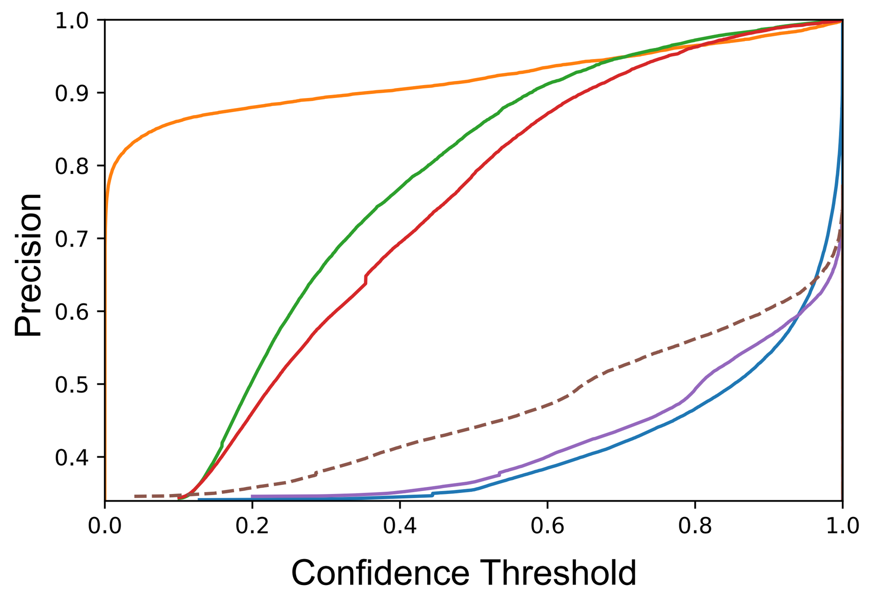


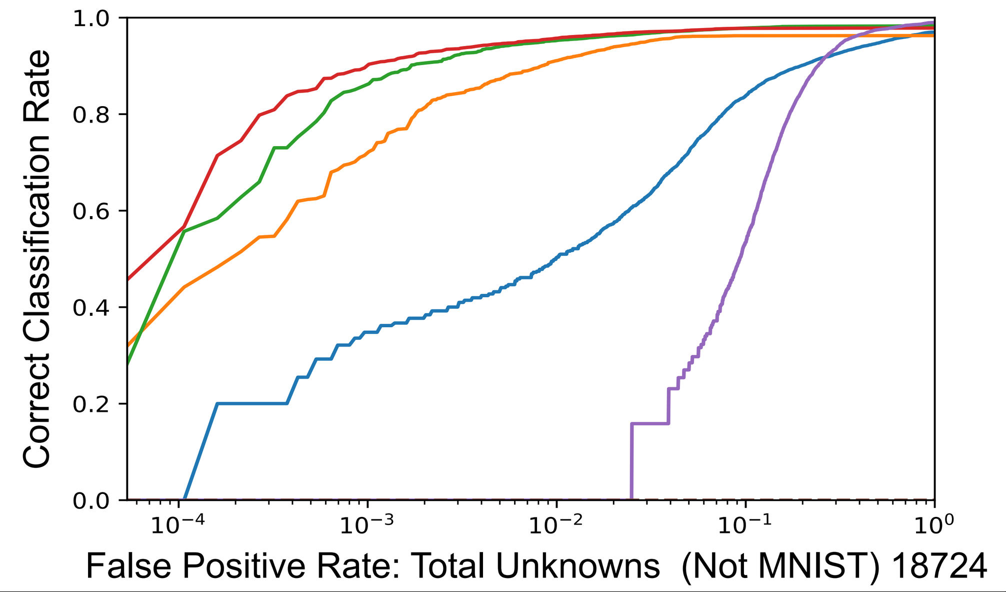
Samples in Openset Problem











Experiment		Algorithm	Correct Classification Rates (CCR) at False Positive Rates (FPR) of			
Architecture D_c D_b	Out of Distribution Unknowns $\|D_a\|$		10^{-4}	10^{-3}	10^{-2}	10^{-1}
LeNet++ D_c → MNIST D_b → NIST Letters	Devanagri 10032	Softmax	0.0	0.0	0.0777	0.9007
		Background	0.0	0.4402	0.7527	0.9313
		Entropic Open-Set	0.7142	0.8746	0.9580	0.9788
		Objectosphere	**0.7350**	**0. 9108**	**0. 9658**	**0. 9791**
	NotMNIST 18724	Softmax	0.0	0.3397	0.4954	0.8288
		Background	0.3806	0.7179	0.9068	0.9624
		Entropic Open-Set	0.4201	0.8578	0.9515	**0.9780**
		Objectosphere	**0.512**	**0.8965**	**0.9563**	0.9773
	CIFAR10 10000	Softmax	0.7684	0.8617	0.9288	0.9641
		Background	0.8232	0.9546	0.9726	0.973
		Entropic Open-Set	**0.973**	**0.9787**	**0.9804**	**0.9806**
		Objectosphere	0.9656	0.9735	0.9785	0.9794
ResNet-18 D_c → CIFAR-10 D_b → CIFAR-100	SVHN 26032	Softmax	0.1924	0.2949	0.4599	0.6473
		Background	0.2012	0.3022	0.4803	0.6981
		Entropic Open-Set	0.1071	0.2338	0.4277	0.6214
		Objectosphere	0.1862	0.3387	0.5074	0.6886
		Scaled Objectosphere	**0.2547**	**0.3896**	**0.5454**	**0.7013**
	CIFAR-100 Subset 4500	Softmax	N/A	0.0706	0.2339	0.5139
		Background	N/A	0.1598	0.3429	0.6049
		Entropic Open-Set	N/A	0.1776	0.3501	0.5855
		Objectosphere	N/A	0.1866	0.3595	0.6345
		Scaled Objectosphere	N/A	**0.2584**	**0.4334**	**0.6647**