Towards Open Set Deep Networks

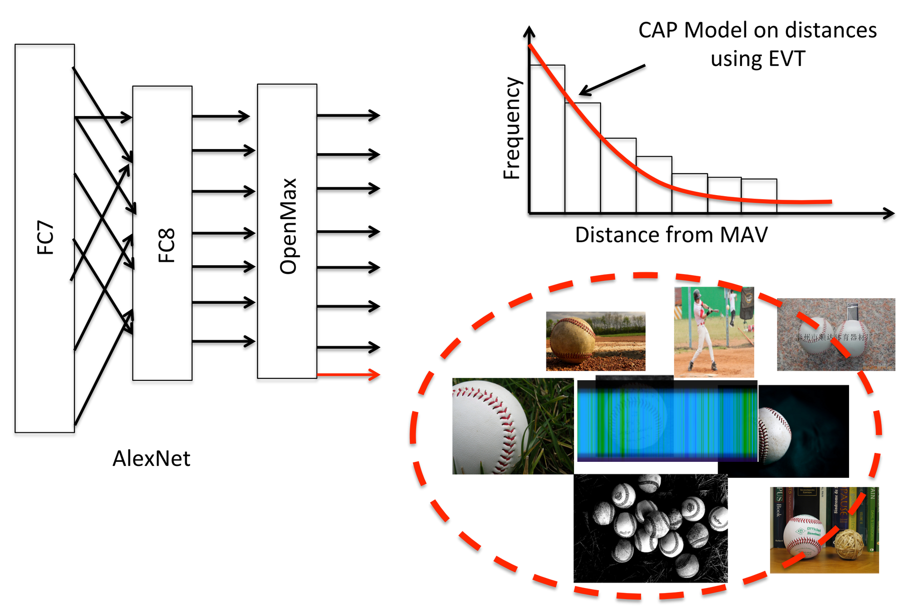
Abhijit Bendale[1,2], Terrance Boult[2]
Samsung Research America[1] and University of Colorado Colorado Springs[2]

CVPR 2016

Contributions

1. Formalization of **Open Set Deep Networks** using **Meta-Recognition** and **OpenMax**.
2. Multi-Class Meta-Recognition using Activation Vectors to estimate the **probability of deep network failure**.
3. Proof that the proposed approach manages **open space risk** for deep networks.
4. Experimental analysis of the effectiveness of open set deep networks at **rejecting unknown classes, fooling images** and obvious errors from adversarial images, while maintaining its accuracy on testing images

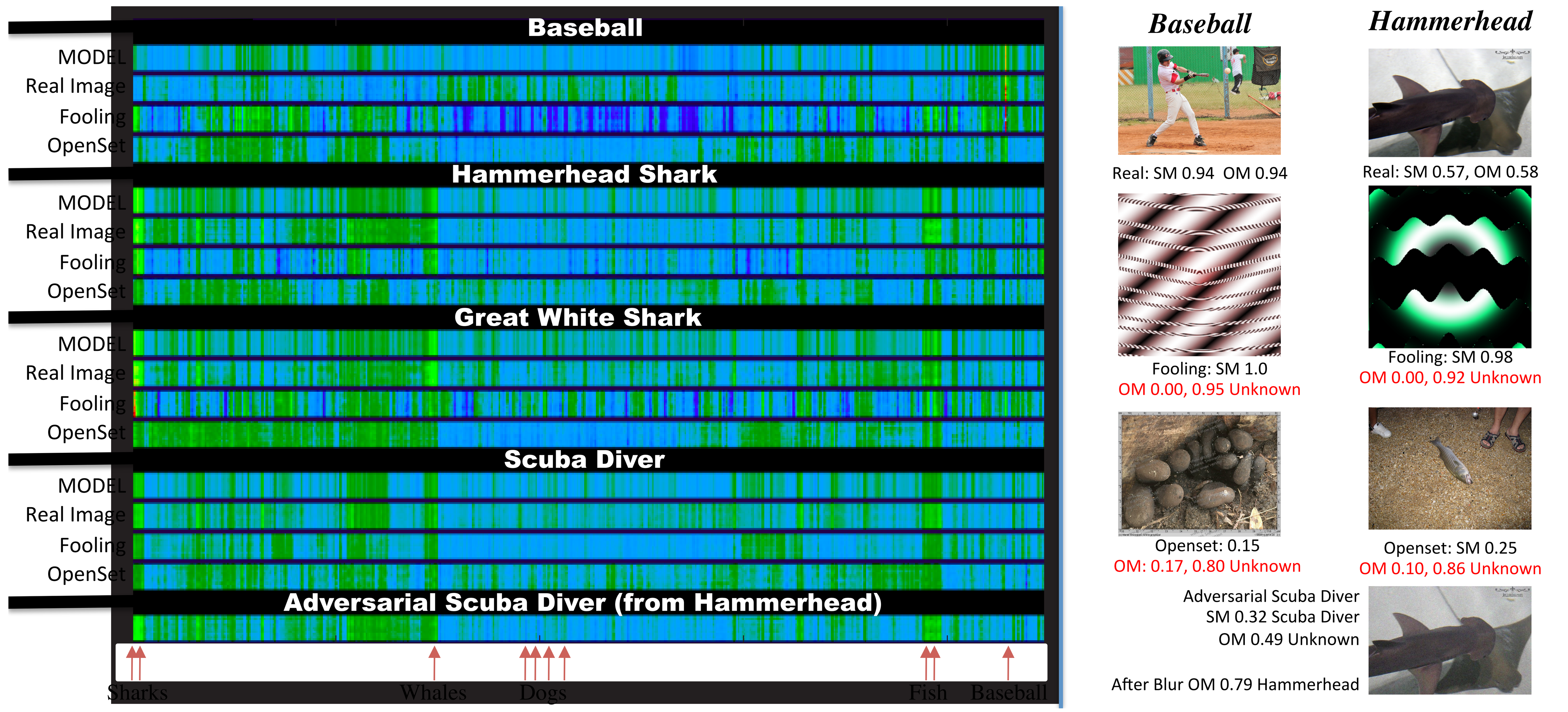
OpenMax Layer for Deep Networks



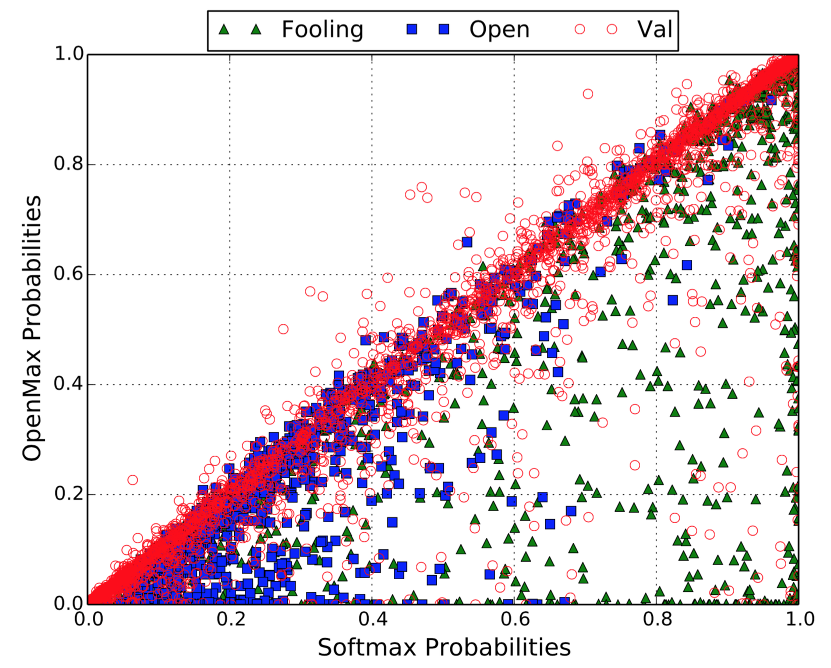
Theorem 1 (Open Set Deep Networks). *A deep network extended using Meta-Recognition on activation vectors as in Alg. 2, with the SoftMax later adapted to OpenMax, as in Eq. 2, provides an open set recognition function.*

Proof. The Meta-Recognition probability (CDF of a Weibull) is a monotonically increasing function of $\|\mu_i - x\|$, and hence $1 - \omega_i(x)$ is monotonically decreasing. Thus, they form the basis for a compact abating probability model. (Please refer the paper for more details about the proof.) □

Mean Activation Vectors

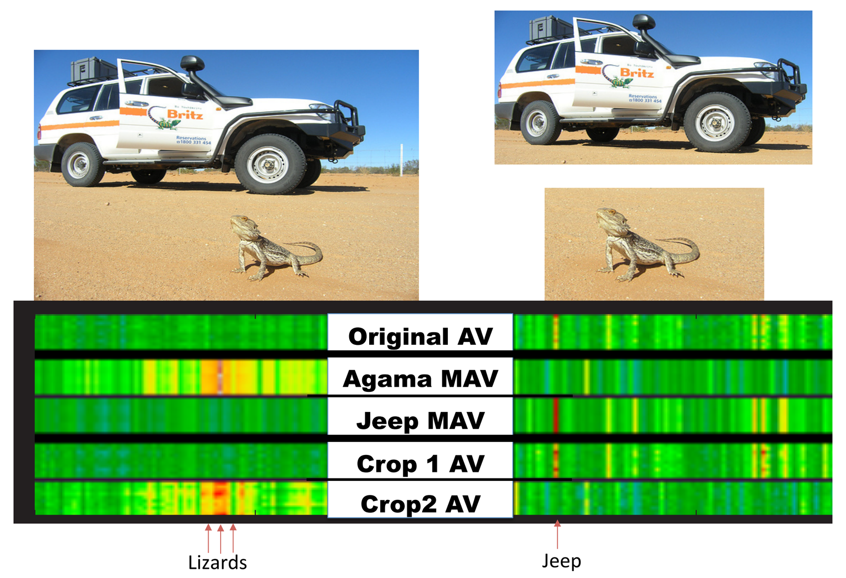


OpenMax Vs SoftMax



1. The more off-diagonal a point, the more OpenMax altered the probabilities
2. For some inputs OpenMax increased the classes probability, which occurs when the leading class is partially rejected thereby reducing its probability and increasing a second or higher ranked class.

Failure Prediction with OpenMax



1. Image from Agama class gets rejected by MAV for Agama. Highest scoring class is jeep with prob 0.26.
2. Crop 1 is the jeep region (accepted with 0.32 probability), crop 2 is agama (accepted with 0.21 probability score).
3. Failure prediction with adversarial images

Meta-Recognition Calibration for OSDN

Algorithm 1 EVT Meta-Recognition Calibration for Open Set Deep Networks, with per class Weibull fit to η largest distance to mean activation vector. Returns libMR models ρ_j which includes parameters τ_i for shifting the data as well as the Weibull shape and scale parameters:κ_i, λ_i.

Require: FitHigh function from libMR
Require: Activation levels in the penultimate network layer $\mathbf{v}(\mathbf{x}) = v_1(x) \dots v_N(x)$
Require: For each class j let $S_{i,j} = v_j(x_{i,j})$ for each correctly classified training example $x_{i,j}$.

for $j = 1 \dots N$ **do**
  Compute mean AV, $\mu_j = mean_i(S_{i,j})$
  EVT Fit $\rho_j = (\tau_j, \kappa_j, \lambda_j) = \text{FitHigh}(\|\hat{S}_j - \mu_j\|, \eta)$
end for
Return means μ_j and libMR models ρ_j

OpenMax

Algorithm 2 OpenMax probability estimation with rejection of unknown or uncertain inputs.

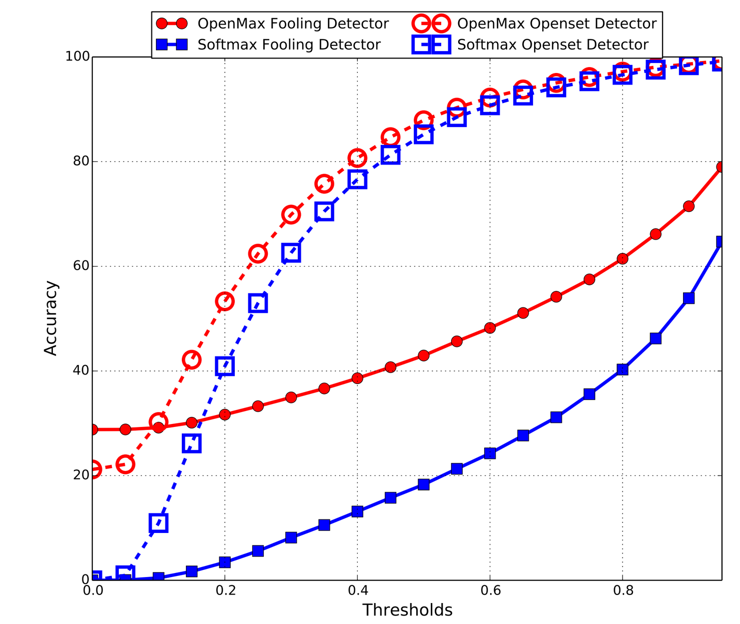
Require: Activation vector for $\mathbf{v}(\mathbf{x}) = v_1(x), \dots, v_N(x)$
Require: **means** μ_j and libMR models $\rho_j = (\tau_i, \lambda_i, \kappa_i)$
Require: α, the numer of "top" classes to revise

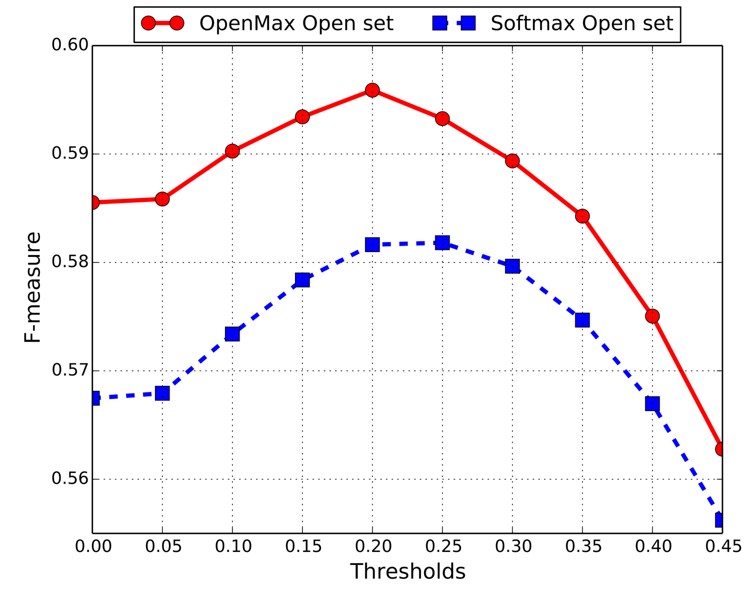
Let $s(i) = \text{argsort}(v_j(x))$; Let $\omega_j = 1$
for $i = 1, \dots, \alpha$ **do**
  $\omega_{s(i)}(x) = 1 - \frac{\alpha - i}{\alpha} e^{-\left(\frac{\|x - \tau_{s(i)}\|}{\lambda_{s(i)}}\right)^{\kappa_{s(i)}}}$
end for
Revise activation vector $\hat{v}(x) = \mathbf{v}(\mathbf{x}) \circ \omega(\mathbf{x})$
Define $\hat{v}_0(x) = \sum_i v_i(x)(1 - \omega_i(x))$.

$$\hat{P}(y = j|\mathbf{x}) = \frac{e^{\hat{\mathbf{v}}_\mathbf{j}(\mathbf{x})}}{\sum_{i=0}^{N} e^{\hat{\mathbf{v}}_\mathbf{i}(\mathbf{x})}} \quad (2)$$

Let $y^* = \text{argmax}_j P(y = j|\mathbf{x})$
Reject input if $y^* == 0$ or $P(y = y^*|\mathbf{x}) < \epsilon$

Experiments





Experiment Details

1. **Dataset: Training** ILSVRC'12: 1.3M training images, 1K classes
2. **Dataset: Testing** 80K images total. 50K Images (1K classes) from ILSVRC'12 Validation set, 15K Fooling Images(1K classes), 15K open set images from 360 classes from ILSVRC'10 (these 360 classes are NOT present in ILSVRC'12).
3. **Model:** BVLC AlexNet (57.1% top-1 accuracy on ILSVRC'12 val set).
4. **Algorithms:** SoftMax, OpenMax and 1-vs-Set algorithm.
5. **Performance:** OpenMax performance gain is nearly 4.3% improvement accuracy over SoftMax with optimal threshold, and 12.3% over the base deep network. Putting that in context, over the test set OpenMax correctly classified **3450 more images than SoftMax with optimal threshold** and **9847 more** than the base deep network. Optimal F-Measure for each algorithm was **SoftMax 0.58**, **OpenMax 0.595** and **1-vs-Set SVM 0.407**.

Acknowledgements: The work support in part by NSF Research Grant IIS-1320956 and UCCS Graduate School Fellowship.
Contact: {abendale,tboult}@vast.uccs.edu
Code/Data: http://vast.uccs.edu/OSDN