Northeastern University
Institute for Intelligent Networked Systems

MENTIS LABORATORY

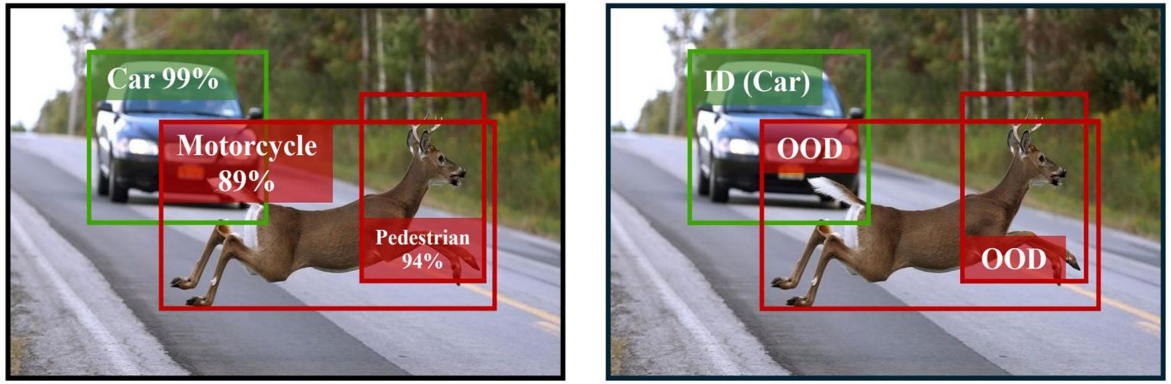
ENCORE: A Neural Collapse Perspective on Out-of-Distribution Detection in Deep Neural Networks

MENTIS laboratory – https://mentis.info/
Northeastern University, United States

Overview

- Importance of OOD Detection
- Current Literature
- Energy Based Score
- Neural Collapse (NC)
- Our Proposal: ENCORE
- Intuition Behind ENCORE
- Key Results
- Takeaways

Importance of OOD Detection



Misdetection of unknown inputs can lead to fatal scenario[1,2]

[1] https://www.gwclaw.com/blog/uber-self-driving-car-death-ntsb-report/

[2] Matos, F., Bernardino, J., Durães, J., & Cunha, J. (2024). A Survey on Sensor Failures in Autonomous Vehicles: Challenges and Solutions. Sensors (Basel, Switzerland), 24(16), 5108. https://doi.org/10.3390/s24165108

Image Source: Zhang, Bin, et al. "Runa: Object-level out-of-distribution detection via regional uncertainty alignment of multimodal representations." AAAI, 2025

Current Literature

Most OOD detectors are:

- Empirical heuristics
- Score-based (MSP[1], Energy[2], GEN[3])
- Feature-space distance or scaling (KNN[4], React[5], ViM[6])

Lack grounding in *established structural properties* of deep networks

When do they fail?

Why do they work?

[1] Hendrycks, Dan, and Kevin Gimpel. "A baseline for detecting misclassified and out-of-distribution examples in neural networks." arXiv preprint arXiv:1610.02136 (2016).

[2] Liu, Weitang, et al. "Energy-based out-of-distribution detection." Advances in neural information processing systems 33 (2020): 21464-21475.

[3] Liu, Xixi, Yaroslava Lochman, and Christopher Zach. "Gen: Pushing the limits of softmax-based out-of-distribution detection." CVPR 2023.

[4] Sun, Yiyou, et al. "Out-of-distribution detection with deep nearest neighbors." International conference on machine learning. PMLR, 2022.

[5] Sun, Yiyou, Chuan Guo, and Yixuan Li. "React: Out-of-distribution detection with rectified activations." Advances in neural information processing systems 34 (2021): 144-157.

[6] Wang, Haoqi, et al. "Vim: Out-of-distribution with virtual-logit matching." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022.

Energy Based Score

$$E(x) = -\log \sum_c \exp(z_c)$$

- z_c are the logits from the DNN
- Affinely connected to data distribution p(**x**) (under heavy assumption)

Limitations

Logits may be spuriously large

Overconfident OOD samples

Energy works theoretically

Lacks in performance

Ignores established DNN properties

Neural Collapse (NC)

Key properties[1]:

(NC1) Features collapse to class means

(NC2) Class means form a simplex ETF

(NC 3)Weights align with class means

(NC4)Classifier reduces to nearest class center

(NC5) OOD features become orthogonal to ID feature space[2]

[1] Papyan, Vardan, X. Y. Han, and David L. Donoho. "Prevalence of neural collapse during the terminal phase of deep learning training." Proceedings of the National Academy of Sciences 117.40 (2020): 24652-24663.

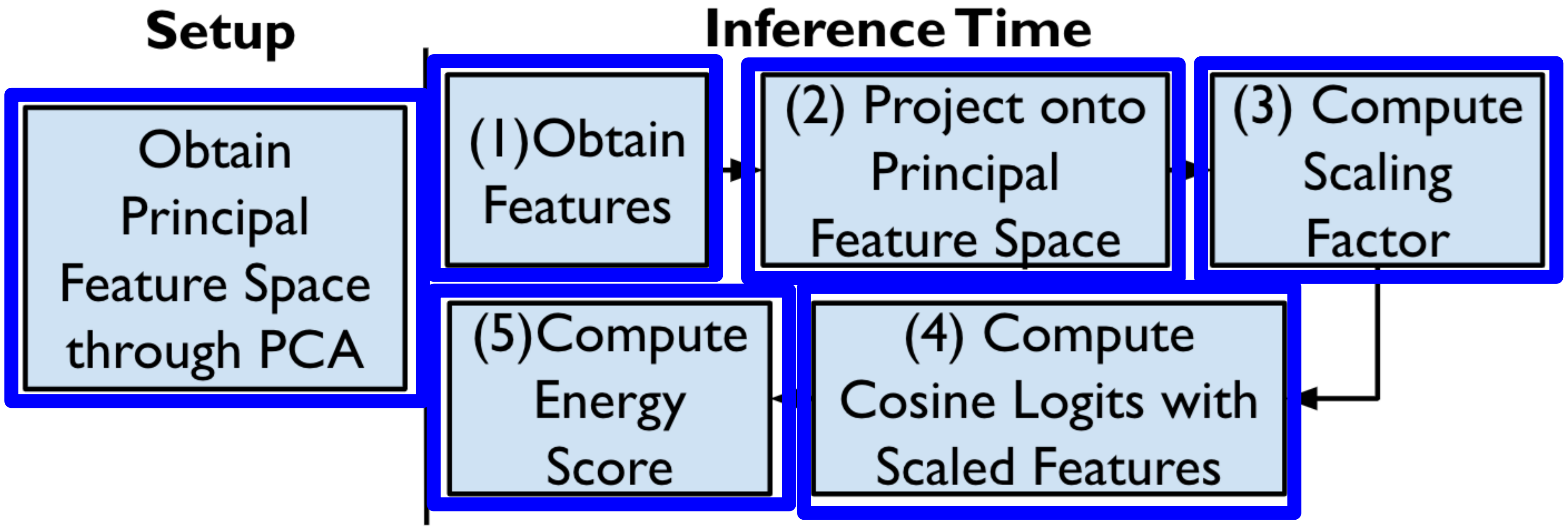
[2] Ammar, Mouïn Ben, et al. "NECO: NEural Collapse Based Out-of-distribution detection." The Twelfth International Conference on Learning Representations.

Our Proposal: ENCORE

- Key insight: Uniform feature scaling improves ID/OOD separation
- With centered feature norm "r", model specific constant "α" and C classes

$$\Delta E \approx \alpha r^2 - \log(C)$$

$$E(\mathbf{x}) = -\log\left[\sum_{i=1}^{C} z_i\right] = -\log\left[\sum_{i=1}^{C} \exp\left(\exp\left(\lambda \frac{\|\mathbf{x}^p\|}{\|\mathbf{x}\|}\right)\cos\theta_i\right)\right]$$



Intuition Behind ENCORE

Why cosine similarity?

- Removes norm bias
- Aligns with NC3 (weights align with class means)

Why adaptive scaling?

- Strong feature alignment with ID subspace should be amplified
- Weakly alignment (likely OOD) should be suppresses

This mimics NC5 – OOD inputs are orthogonal to ID feature space

Key Results

	Near OOD		Far OOD	
	FPR	AURC	FPR	AURC
Energy	68.18	86.95	40.41	91.80
React	71.07	86.51	42.08	91.09
ViM	47.69	88.55	25.68	93.17
GEN	57.89	87.80	33.59	91.57
ENCORE	**36.27**	**91.23**	**21.11**	**94.56**

Comparative Results for CIFAR10 Benchmark

Key Results

	Near OOD		Far OOD	
	FPR	AUROC	FPR	AUROC
React	75.79	72.61	71.11	77.55
ViM	70.13	**78.66**	**28.36**	90.71
GEN	67.08	77.33	29.82	90.08
ENCORE	**66.01**	77.70	28.53	90.72

Comparative Results for Imagenet Benchmark

Key Results

Table 8. Comparison of ENCORE with similar approaches on the ImageNet benchmark using ViT-Huge-14, showing results for Near-OOD and Far-OOD datasets.

	Near OOD		Far OOD	
	FPR ↓	AUROC ↑	FPR ↓	AUROC ↑
SCALE	58.24	77.52	26.53	88.51
NECO	55.55	79.64	24.25	89.96
ENCORE (Ours)	**43.55**	**89.47**	**14.62**	**96.89**

Table 3. Inference latency (ms) for different OOD detection methods on RepVGG-a2 with batch size 1024.

Method	Latency (ms)
ENCORE	135
ViM	155
KNN	354

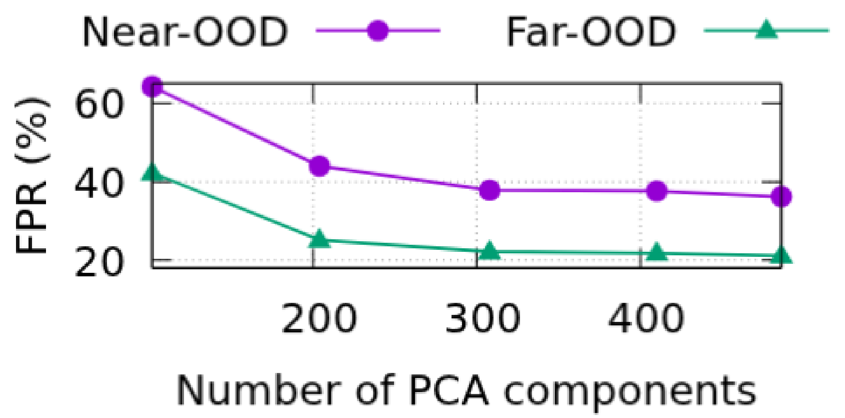


Figure 3. Variation of FPR with varying number of dimensions for principal feature-space. The results are for ResNet18 and CIFAR10 benchmark.

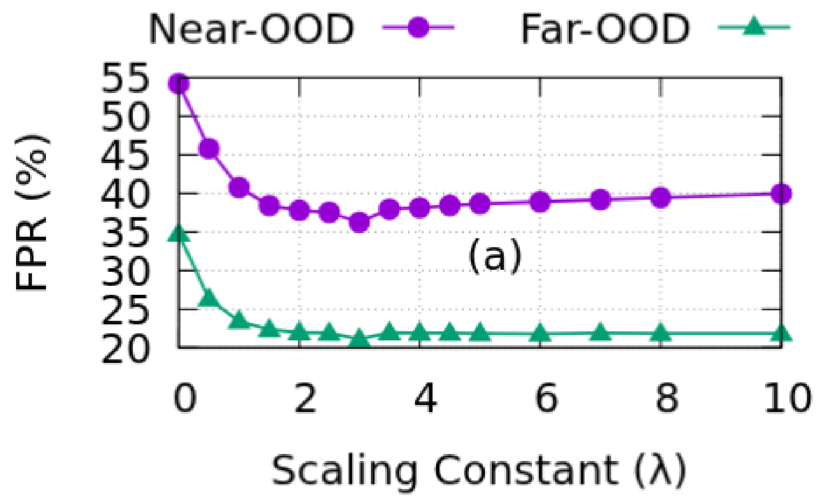


Figure 4. Variation of FPR with varying the scaling constant (λ)

Takeaways

Neural Collapse give theoretical basis for improvement of OOD detection

Uniformly scaling the features improves ID/OOD separation

Adaptive scaling provides further gain in performance

Result is contingent upon agreement to NC theory

Northeastern University
Institute for Intelligent
Networked Systems

MENTIS
LABORATORY

Thank you!

MENTIS laboratory – https://mentis.info/
Northeastern University, United States