Towards Deep Anomaly Detection with Structured Knowledge Representations

(Doctoral Track)

Konstantin Kirchheim

Deptartment of Computer Science
Otto-von-Guericke University Magdeburg
Germany

September 17, 2023

Agenda

Motivation

Research Objectives

Proposed Solution

Proof of Concept

Conclusion

Motivation

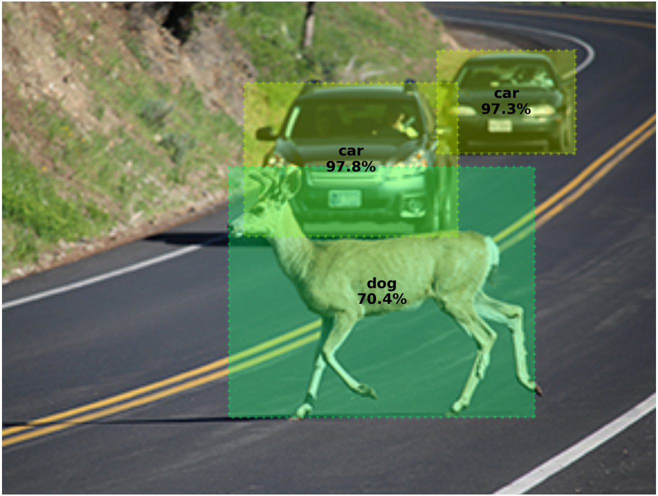

Figure: Deer (Unknown Class) misclassified

Motivation

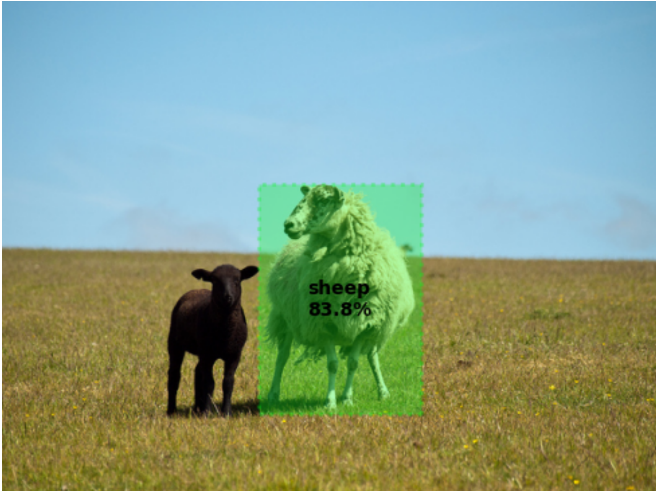

Figure: Sheep (Known Class) not detected

Motivation

Figure: Moose (Unknown Class) not detected

Deep Anomaly Detection

Anomaly Detection with Deep Neural Networks (DNNs)

- Out-of-Distribution Detection [1]
- Outlier Detection [2]
- Novelty Detection [3]

Autonomous Agents in Open Environments

- Have hypotheses about the world
- Example: *"All stop signs are red"*
- Violations - anomalies - are potentially safety-critical
  $\rightarrow$ Should be detected

Limitations of Current Deep Methods

- Not explainable
  $\rightarrow$ Only provide score
  $\rightarrow$ Has model actually learned what we want it to?
- Integrating prior knowledge is not straight-forward
  $\rightarrow$ Lots of data required to learn simple concepts
- Arguably no robust high-level reasoning

Constructing a problem that current methods can not solve is surprisingly simple...

SuMNIST

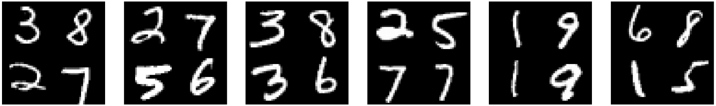

Figure: Sample of SuMNIST

Method	Backbone	AUROC ↑	AUPR-IN ↑	AUPR-OUT ↑	FPR95 ↓
Nearest Neighbor	-	50.00	59.18	90.82	100.00
Deep Nearest Neighbor [4]	ViT-L/16	51.19	18.81	82.21	94.34
Mahalanobis	-	50.00	59.18	82.31	100.00
Mahalanobis [5]	ViT-L/16	50.00	59.18	84.58	100.00
Deep SVDD [6]	-	49.32	18.07	81.28	95.14

Table: SOTA is close to random guessing

Could reasoning emerge from scaling?

Maybe, but:

- Studies on OOD detection: models reached limit (data/computation) [7]
- Scaling Transformers seems ineffective for reasoning tasks [8]

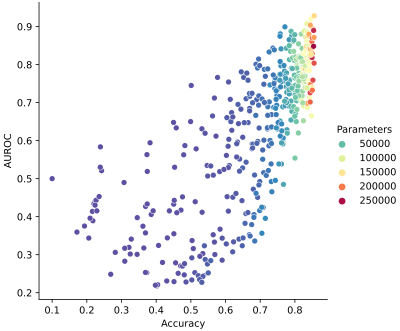

Figure: OOD Detection on CIFAR10

Hypothesis

Problem:

- World-Knowledge in DNNs is not structured

Using structured knowledge representations

- allows to integrate priors about structure of $p(x)$
- improves the robustness
- improves data efficiency
- increases explainability

Research Objectives

Classification

- *"All German stop signs are red octagons."*

Object Detection

- *"A human face is part of a human."*

Temporal Dynamic

- *"There is a limit to the velocity of objects."*

Structure Learning

- Can we learn structure directly from the data?

Structured Knowledge Representation

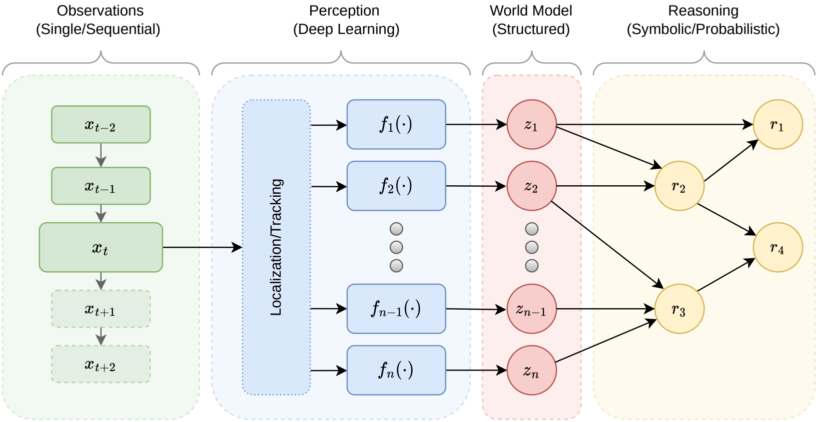

Figure: Proposed Architecture

Proof of Concept

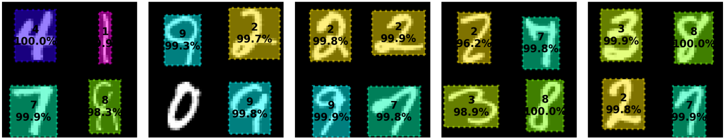

Figure: Sample of SuMNIST with Detected Objects

Hybrid Class:

- Saves all combinations of numbers observed during training
- Does not require class $\rightarrow$ number mapping

Hybrid Sum:

- Calculate sum of all detected numbers
- Requires class $\rightarrow$ number mapping

Experiments

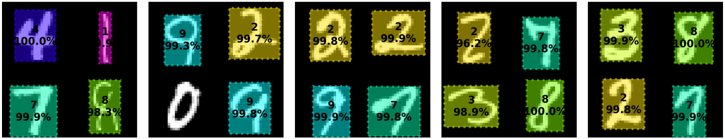

Figure: Sample of SuMNIST with Detected Objects

Method	Backbone	AUROC ↑	AUPR-IN ↑	AUPR-OUT ↑	FPR95 ↓
Nearest Neighbor	-	50.00	59.18	90.82	100.00
Deep Nearest Neighbor [4]	ViT-L/16	51.19	18.81	82.21	94.34
Mahalanobis	-	50.00	59.18	82.31	100.00
Mahalanobis [5]	ViT-L/16	50.00	59.18	84.58	100.00
Deep SVDD [6]	-	49.32	18.07	81.28	95.14
Hybrid Memory (**Ours**)	ResNet-18	95.30	82.72	99.29	9.26
Hybrid Sum (**Ours**)	ResNet-18	**98.41**	**92.69**	**99.76**	**2.98**

Table: Results

Conclusion

- Current models can not reason robustly
- Scaling might not fix this
- We propose framework to address this
- Can outperform SOTA

Future Work:

- Large Datasets
- Videos
- Structure Learning

Figure: GitHub

References I

Jingkang Yang, Kaiyang Zhou, Yixuan Li, and Ziwei Liu.
Generalized out-of-distribution detection: A survey.
ArXiv, abs/2110.11334, 2021.

Charu C Aggarwal.
Outlier Analysis.
Springer, 2017.

Stanislav Pidhorskyi, Ranya Almohsen, Donald A Adjeroh, and Gianfranco Doretto.
Generative probabilistic novelty detection with adversarial autoencoders.
Advances in Neural Information Processing Systems, 31, 2018.

Liron Bergman, Niv Cohen, and Yedid Hoshen.
Deep nearest neighbor anomaly detection.
arXiv preprint arXiv:2002.10445, 2020.

References II

Kimin Lee, Kibok Lee, Honglak Lee, and Jinwoo Shin.
A simple unified framework for detecting out-of-distribution samples and adversarial attacks.
Advances in Neural Information Processing Systems, 31, 2018.

Lukas Ruff, Robert Vandermeulen, Nico Goernitz, Lucas Deecke, Shoaib Ahmed Siddiqui, Alexander Binder, Emmanuel Müller, and Marius Kloft.
Deep one-class classification.
In International conference on machine learning, pages 4393–4402. PMLR, 2018.

References III

- Dan Hendrycks, Steven Basart, Mantas Mazeika, Mohammadreza Mostajabi, Jacob Steinhardt, and Dawn Song.
  Scaling out-of-distribution detection for real-world settings.
  arXiv preprint arXiv:1911.11132, 2019.

- Dan Hendrycks, Collin Burns, Saurav Kadavath, Akul Arora, Steven Basart, Eric Tang, Dawn Song, and Jacob Steinhardt.
  Measuring mathematical problem solving with the math dataset.
  arXiv preprint arXiv:2103.03874, 2021.