Advancing Out-of-Distribution Detection via Local Neuroplasticity

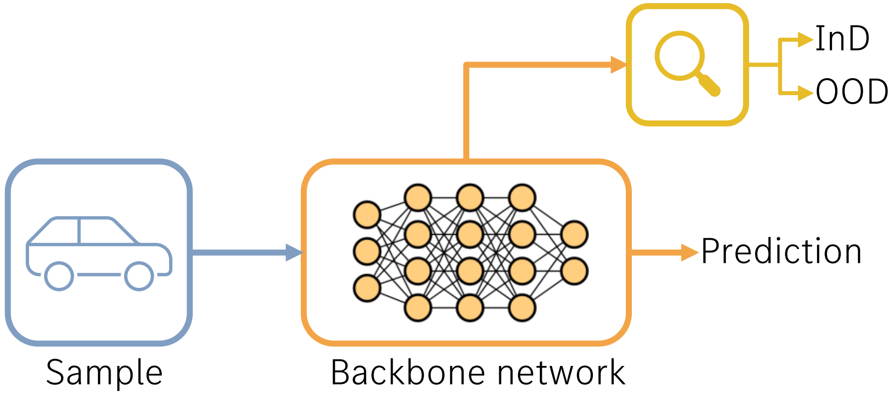
Alessandro Canevaro[*,1,2], Julian Schmidt[1], Mohammad Sajad Marvi[1], Hang Yu[1], Georg Martius[2], Julian Jordan[1]

[1]Mercedes-Benz AG, Sindelfingen, Germany [2]University of Tübingen, Tübingen, Germany

*alessandro.canevaro@mercedes-benz.com

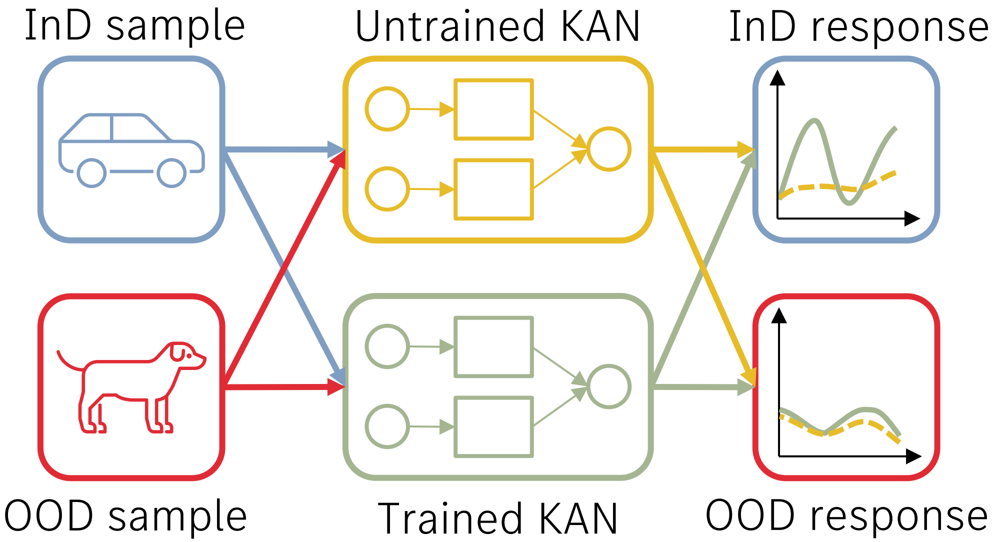
Out-Of-Distribution (OOD) Detection

- Machine learning models often assume that training and test data share the same distribution, yet real-world scenarios frequently expose models to samples that differ significantly from the training set.
- OOD detection addresses this challenge by identifying inputs that deviate from the expected in-distribution data, thereby enhancing model robustness.
- OOD detection methods operate on the latent space of pre-trained backbone models, where they capture semantic shifts between In-Distribution (InD) and OOD samples.
- OOD detection remains a challenging task as OOD types can vary widely, and current approaches often struggle to maintain high performance across different datasets.



Proposed Approach

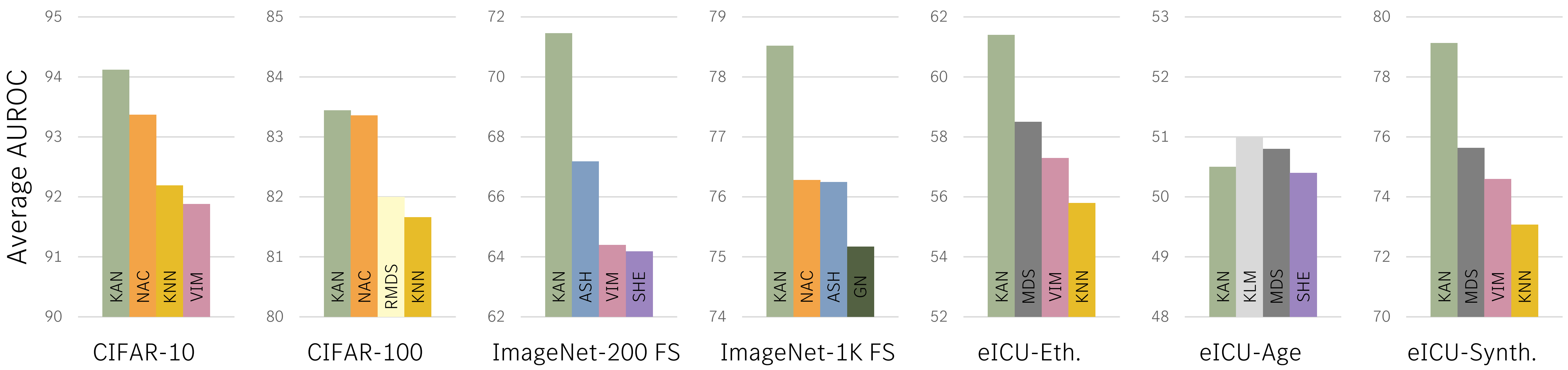
- Kolmogorov-Arnold Networks (KANs) are neural architectures based on the Kolmogorov-Arnold theorem that decompose high-dimensional functions into sums of univariate functions.
- Local neuroplasticity means only the spline coefficients activated by the InD samples are adjusted during training, preserving other regions.
- We exploit this by comparing activation patterns between a trained and untrained KAN. During inference (detection) time, if the sample is OOD, both networks will have the same response; otherwise, there will be a difference in the responses indicating that the sample is InD.



Mercedes-Benz | EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Results

- The experimental results show that our KAN-based detector reaches SOTA performance across seven benchmarks from two different domains.



*Detailed results with per-dataset metrics and comparison with all 22 baseline methods are available in the paper.

Conclusion

- Our method outperforms SOTA approaches across diverse benchmarks, in both image and tabular data domains.
- Ablation studies show its robustness to parameter variations and training dataset size.

Acknowledgments

This work is a result of the joint research project STADT:up (19A22006O). The project is supported by the German Federal Ministry for Economic Affairs and Climate Action (BMWK), based on a decision of the German Bundestag. The authors are solely responsible for the content of this publication.